**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/02, 15/52, A61K 35/00 — A1

(11) International Publication Number: **WO 00/47725**

(43) International Publication Date: 17 August 2000 (17.08.00)

(21) International Application Number: PCT/GB00/00431

(22) International Filing Date: 10 February 2000 (10.02.00)

(30) Priority Data:
9903019.9 10 February 1999 (10.02.99) GB

(71) Applicant *(for all designated States except US)*: MICROBIOLOGICAL RESEARCH AUTHORITY [GB/GB]; CAMR, Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MINTON, Nigel [GB/GB]; Microbiological Research Authority, CAMR, Porton Down, Salisbury, Wiltshire SP4 0JG (GB). ANLEZARK, Gill [GB/GB]; Microbiological Research Authority, CAMR, Porton Down, Salisbury, Wiltshire SP4 0JG (GB). VAUGHAN, Thomas [GB/GB]; Department of Biology, University of York, P.O. Box 373, York YO10 5YW (GB).

(74) Agent: HARDING, Charles, D.; D. Young & Co., 21 London Road, Southampton SO15 2AD (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: NITROREDUCTASE ENZYMES

(57) Abstract

The present invention relates to polypeptides and proteins having nitroreductase activity. The invention also relates to DNA and genes encoding these nitroreductases, and to methods of obtaining such enzymes, DNA and genes. In a particularly preferred aspect, the nitroreductase enzymes demonstrate preferential catalytic conversion of the alkylating agent CB1954 into its highly cytotoxic 4–hydroxylamine (4HX) derivative, this derivative demonstrating anticarcinoma properties. Accordingly, the catalytic activity of the nitroreductase enzymes of the present invention may be employed to achieve catalysis of CB1954 into its cytotoxic derivative in a site–directed manner, such as by Directed–Enzyme Prodrug Therapy (DEPT).

B5

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# NITROREDUCTASE ENZYMES

The present invention relates to polypeptides and proteins having nitroreductase activity, to DNA and genes encoding these nitroreductases and to methods of obtaining such enzymes, DNA and genes.

A number of cancer therapies are based upon or exploit the conversion of a non-toxic prodrug into a toxic derivative.

One example concerns the monofunctional alkylating agent CB1954, which exhibits extreme toxicity towards the Walker 256 rat carcinoma as a result of the presence of a DT-diaphorase enzyme (DTD) which reduces the 4-nitro group of CB1954 to give a highly cytotoxic 4-hydroxylamine (4HX) derivative. CB1954 does not have the same effect on human carcinomas because human cells lack this enzyme but would be effective against human tumours if an enzyme such as DTD were externally supplied, e.g. in a Directed-Enzyme Prodrug Therapy (DEPT). The rat DTD, however, has a relatively poor specific activity for CB1954. The *E.coli* B nitroreductase enzyme (NfnB) was isolated as a more effective alternative and is the subject of EP-A-0540263. It exhibits a higher specific activity for CB1954, compared with the rat enzyme and is, therefore, currently the preferred enzyme in anti-cancer DEPT strategies.

Whilst the known *E.coli* enzyme receives widespread attention from cancer biologists seeking to develop gene based DEPT strategies, it has a number of drawbacks. These mostly relate to its activity against the preferred prodrug, CB1954 - it has a relatively high $K_m$ and low $K_{cat}$, and converts CB1954 into equimolar amounts of a relatively innocuous 2-hydroxylamino derivative (2HX) in addition to the highly cytotoxic 4-hydroxylamino species (4HX).

In relation to this specific prodrug, it is hence desired to provide an

alternative to the known *E.coli* enzyme.

Additionally, and more generally, analogues of CB1954 and prodrugs other than CB1954 are known and further such precursors of potential toxic agents may become the focus of future therapies. In relation to all of these it is desired to provide further enzymes capable of use in converting prodrugs into drugs, e.g. for clinical uses.

It is an object of the present invention to provide nitroreductase enzymes, in particular nitroreductase enzymes for converting CB1954 and analogues thereof into drugs. It is a further object of the present invention to provide DNA and genes encoding nitroreductases, which DNA and genes in particular are incorporated into pharmaceutical compositions for prodrug therapies.

The present invention is based upon the discovery, purification, gene sequencing and/or expression of nitroreductases in bacteria and other microorganisms with hitherto unknown properties in converting prodrugs such as CB1954 into toxic derivatives. These nitroreductases posses properties which alone or in combination offer potential improvements compared with the known enzymes in this technology. The nitroreductases of the invention may be divided into different families based upon such characteristics as activity, product spectrum and/or amino acid sequence, and each given nitroreductase may fall into more than one of these families.

The present invention provides, in a first aspect, a nitroreductase enzyme, characterised in that it preferentially reduces CB1954 to a product that is a cytotoxic 4-hydroxylamine (4HX) derivative.

The enzymes of this aspect of the present invention confer the advantage that the product they generate from CB1954 contains a greater proportion

of the cytotoxic 4HX derivative then the non-cytotoxic 2-hydroxylamino derivative. In preferred embodiments of the invention, the product is substantially entirely the cytotoxic derivative. The enzymes may hence be more efficient that those of the art as the enzymes of the invention produce more cytotoxic product for a given amount of pro-drug.

The present invention further provides, in a second aspect, a nitroreductase enzyme, characterised in that it reduces a prodrug to a toxic derivative with a $K_m$ of less 700 micromolar, wherein the prodrug is selected from CB1954 and analogues thereof or other bioreductive drugs (Denny et al, B.J. Cancer, 1996, 74, pp S32-S38). The enzymes of the second aspect of the invention offer an advantage over the known *E.coli* - derived enzyme in that they have a lower $K_m$ ($K_m$ of *E.coli* NfnB for CB1954 is around 862 micromolar) and thus have a higher affinity for substrate. Twenty nitrogen mustard analogues of CB1954 are described by Friedlos et al (J Med Chem, 1997, 40, 1270-1275).

More preferably, the $K_m$ of the enzymes of the second aspect of the invention is less than 300 micromolar.

In a third aspect, the present invention provides a nitroreductase enzyme characterised in that it reduces a prodrug to a toxic derivative with a $K_{cat}$ of at least 8, wherein the prodrug is selected from CB1954 and analogues thereof.

The enzymes of this aspect of the invention offer an improvement over that of the art, specifically the *E.coli* enzyme, in that they have an improved $K_{cat}$ - i.e a higher value than for *E.coli* NfnB indicating a higher turnover of substrate by the enzyme. In preferred embodiments of this aspect of the invention, the $K_{cat}$ of the enzymes is at least 10.

In a fourth aspect of the invention, there is provided a nitroreductase

enzyme characterised in that it reduces CB1954 to a toxic derivative, it reduces SN23862 to a toxic derivative, it can use NADH and/or NADPH as electron donor and in that it shares no more than 50% sequence identity with the *E.coli* NfnB sequence. Preferably, the sequence identity is about 25% or less, this sequence identity being measured using the MEGALIGN (registered trade mark) software.

It has already been discussed how the known *E.coli* nitroreductase is well characterised and is fully sequenced. The nitroreductases of the fourth aspect thus represent a class of enzymes having nitroreductase activity, or being nitroreductase-like, which nevertheless are so different in amino acid sequence from the *E.coli* enzyme as to represent a separate family of nitroreductases.

This aspect of the invention thus advantageously provides a further class of nitroreductase enzymes for use e.g. in prodrug therapies.

The invention still further provides, in a fifth aspect, a nitroreductase enzyme characterised in that it reduces CB1954 or an analogue thereof to a toxic derivative, in that it shares at least 50% sequence identity with the rat DTD sequence and in that it does not contain a domain that is the same as or corresponds to amino acids 51 to 82 of the rat DTD sequence.

Sequence identity is suitably measured in the same way as described above in relation to the fourth aspect.

To determine whether a given nitroreductase contains a domain that is the same as or corresponds to amino acids 51 to 82 of the rat DTD sequence, the amino acid sequence of the given nitroreductase and of the rat DTD sequence are aligned using a conventional sequence alignment program, such as MEGALIGN (registered trade mark) made by DNASTAR, Inc.

If the alignment program indicates that there are no amino acids in the given sequence that, following the algorhythm of the program, are held to correspond to those at positions 51-82 of the rat DTD sequence then it is concluded that the rat domain is lacking from the given sequence.

This aspect of the invention thus provides a further class of nitroreductase enzymes for conversion e.g. of prodrugs into drugs. A nitroreductase in this class may also be obtained by deleting amino acid residues that correspond to residues 51-82 of the rat DTD from a known mammalian enzyme.

The nitroreductases of the invention may also be NADPH dependant. This property further distinguishes some enzymes of the invention from the known *E.coli* enzyme and the rat DTD.

It has been found that enzymes having one or more of the properties described may be obtained from bacteria of the family *Bacillus*, in particular a *Bacillus* selected from *B. amyloliquefaciens*, *B. subtilis*, *B. pumilis*, *B. lautus*, *B. thermoflavus*, *B. licheniformis* and *B. alkophilus*. This finding is of surprise in that at least three nitroreductase enzymes have been found in some species, in particular *B.subtilis*, *B.lautus* and *B.pumilis*, and as nitroreductases having the advantageous properties of the invention have not hitherto been identified in these bacteria, the currently used nitroreductase being obtained from *E.coli*.

In specific embodiments of the invention described in more detail below, a nitroreductase has a sequence selected from SEQ ID Nos 2, 4, 6, 8, 10, 12, 14, 16, 17, 18, 19, 20, 21, 23, 25, 27 and 29.

It has further been found that nitroreductases according to the invention may fall into more than one aspects of the invention. It is hence preferred that a nitroreductase of the invention possesses the properties of at least

two aspects of the invention, and more preferably at least three aspects of the invention.

A specific embodiment of the invention is a nitroreductase of SEQ ID NO:2 obtained from *B. amyloliquefaciens* this enzyme converts CD194 into substantially only the cytotoxic derivative, hence falling into the first aspect of the invention, but also has a $K_m$ that is improved compared to the *E.coli* enzyme, hence falling also into the second aspect of the invention.

A further specific embodiment of the invention is a nitroreductase from *B.subtilis*, SEQ ID NO:9. This enzyme has a better $K_{cat}$ than the *E.coli* enzyme, its $K_{cat}$ being about 15 compared with about 6 for the E.coli enzyme, and hence falls into the third aspect of the invention. Additionally, this enzyme falls into the fourth aspect of the invention in that it reduces both CB1954 and SN23862 but shares less than 30% sequence identity with the *E.coli* sequence. Another *B.subtilis* enzyme, SEQ ID NO:11 is similarly in both the third and fourth aspects of the invention, having a $K_{cat}$ of about 15.

From the examples set out below it will be apparent how the further specific embodiments of the invention fall into at least two and even three aspects of the invention.

The enzymes of the invention are of use in enzyme directed prodrug therapy. Accordingly, it is preferred that they are provided in purified form.

A sixth aspect of the invention provides a pharmaceutical composition comprising a nitroreductase enzyme according to any of the first to fifth aspects of the invention in combination with a pharmaceutically acceptable carrier.

As mentioned above, the nitroreductase of the invention are of use in

therapies such as directed-enzyme prodrug therapies. In these therapies, it is required to deliver the nitroreductase to the target site. This delivery can be achieved by delivering the enzyme itself or by delivering a DNA or gene coding for the enzyme.

In an example of the enzyme of the invention in use, a pharmaceutical composition is designed for a directed-enzyme prodrug therapy, and comprises a pharmaceutically acceptable carrier and a compound for converting a prodrug into a drug, wherein a compound is composed of at least a nitroreductase according to any of the first to fifth aspects of the invention conjugated to a targeting moiety.

The targeting moiety can suitably comprise an antibody specific for a target cell. Alternatively, the targeting moiety is a moiety preferentially accumulated by or taken up by a target cell.

A further example of delivery of the enzyme of the invention is achieved in a gene therapy-based approach for targeting cancer cells, as described in WO 95/12678. As described by Knox R.J. et al, the basis of this further prodrug therapy is delivery of a drug susceptibility gene into target, usually tumour or cancer, cells. The gene encodes a nitroreductase that catalyses the conversion of a prodrug into a cytotoxic derivative. The nitroreductase itself is not toxic and cytotoxicity used to treat the tumour cells arises after administration of a prodrug which is converted into the cytotoxic form. A bystander effect may be observed as cytotoxic drug may diffuse into neighbouring cells.

Thus, in this gene-based therapy, the nitroreductase is expressed inside a cell, in contrast to other delivery systems in which, for example, the enzyme itself is delivered accompanied by a targeting moiety.

Targeting of gene-based therapies may be achieved by providing a virus or

liposome with altered surface components so that the delivery vehicle is recognised by target cells. Typically, transcriptional elements are chosen so that the gene coding for the nitroreductase enzyme will be expressed in the target cells, and preferably substantially only in the target cells. A number of viral-based vectors are suitable for this delivery. Retro-viral based vectors typically infect replicating cells. Adenoviral vectors and lentiviral-vectors are also believed to be suitable.

This delivery technology has been demonstrated by Bridgewater et al (Eur J Cancer 31a, 236-2370, 1995). A recombinant retrovirus encoding a nitroreductase was used to infect mammalian cells, it being observed that infected cells expressing the nitroreductase were killed by application of CB1954.

Accordingly, a further aspect of the invention provides the use of a DNA sequence coding for a nitroreductase of the invention in manufacture of a medicament for prodrug therapy.

The medicament may take the form of a viral vector, comprising a DNA encoding the nitroreductase of the invention operatively coupled to a promoter for expression of the DNA. The medicament may take the form of a mini-gene comprising a DNA operatively linked to a promoter for expression of the DNA, the mini-gene being suitable for inclusion or incorporation into a targeting vehicle such as a microparticle.

Thus, an embodiment of the invention provides a viral vector comprising a nucleotide sequence encoding a nitroreductase according to any of aspects 1 to 5 of the invention, which nitroreductase converts a prodrug into a cytotoxic drug, and also a kit comprising the viral vector and the prodrug, and also a method of treatment of tumours which comprises administering an effective amount of the viral vector together with an effective amount of the prodrug.

The preparation and administration of these viral vectors may be substantially as described in WO 95/12678, the contents of which is incorporated herein by reference. The present invention relates to providing nitroreductase enzymes and genes and DNA coding therefore. The uses of those enzymes and genes may be as set out in WO 95/12678.

A nitroreductase can also be delivered by putting a gene of the invention into a bacteria that selectively colonises tumours, such as a clostridial (Lemmon et al, Gene Therapy, 1997, 4, 791-796) or Salmonella species.

A further aspect of the invention provides an isolated DNA encoding a nitroreductase according to any of the first to fifth aspects of the invention. The DNAs of this further aspect of the invention, and also the DNAs incorporated into vectors of the invention, preferably comprise a sequence which is selected from SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 22, 24, 26 or 28, together with fragments, derivatives and analogs thereof retaining nitroreductase activity according to one of the first to fifth aspects of the invention. The fragments, derivatives and analogs are suitably selected from sequences which retain at least 70% identity with the specific embodiments of the invention, or preferably at least 90% identity and most preferably at least 95% identity.

The enzymes of the invention can also be obtained by purification from cell extracts and may also be obtained by recombinant expression of DNA. A still further aspect of the invention lies in a method of preparing a nitroreductase enzyme, comprising expressing a gene in a bacterial cell, wherein the gene codes for a nitroreductase enzyme of the invention.

In an example of the invention described below in more detail, the gene expressed is a *Bacillus* gene or is a gene obtained by substitution, deletion and/or addition of nucleotides in or to a *Bacillus* gene.

The invention also provides the use of a nitroreductase according to any of the aspects of the invention in manufacture of a medicament for anti-tumour therapy, and the use of a compound comprising a nitroreductase according to any aspect of the invention conjugated to a targeting moiety in manufacture of a medicament for anti-tumour therapy.

The invention is now illustrated by the following specific examples and in the accompanying sequence listing in which:

SEQ ID NO: 2 is a nitroreductase from *B.amyloliquefaciens* (coded for by SEQ ID NO: 1) and designated "Bam YrwO";

SEQ ID NO: 4 is a nitroreductase from *B.subtilis* (coded for by SEQ ID NO: 3) and designated "Bs YwrO";

SEQ ID NO: 6 is a nitroreductase from *B.subtilis* (coded for by SEQ ID NO: 5) and designated "YrkL";

SEQ ID NO: 8 is a nitroreductase from *B.subtilis* (coded for by SEQ ID NO: 7) and designated "YdeQ";

SEQ ID NO: 10 is a nitroreductase from *B.subtilis* (coded for by SEQ ID NO: 9) and designated "Ydgl";

SEQ ID NO: 12 is a nitroreductase from *B.subtilis* (coded for by SEQ ID NO: 11) and designated "YodC";

SEQ ID NO: 14 is a nitroreductase from *E.coli* (coded for by SEQ ID NO: 13) and designated "YabF"

SEQ ID NO: 16 is a nitroreductase from *E.coli* (coded for by SEQ ID NO: 15) and designated "YheR";

SEQ ID NO: 17 is a nitroreductase from *H.influenzae*;

SEQ ID NO: 18 is a nitroreductase from *T.aquaticus*;

SEQ ID NO: 19 is a nitroreductase from *Synechocystis sp* PCC 6803;

SEQ ID NO: 20 is a nitroreductase from *A.fulgidus*;

SEQ ID NO: 21 is a nitroreductase from *A.fulgidus*.

SEQ ID NO: 23 is a nitroreductase from *Campylobacter jejuni* (coded for by SEQ ID NO: 22);

SEQ ID NO: 25 is a nitroreductase from *Porphyromonas gingivalis*

(coded for by SEQ ID NO: 24);

SEQ ID NO: 27 is a nitroreductase from *Yersinia pestis* (coded for by SEQ ID NO: 26); and

SEQ ID NO: 29 is a nitroreductase from Helicobacter pylori (coded for by SEQ ID NO: 28).

The invention is also illustrated by reference to the accompanying Tables 1-4 and Figures 1 and 2, in which Figs 1 and 2 show sequence comparisons as set out in more detail in Example 8.

**Example 1**

**A Nitroreductase Enzyme/Gene from *Bacillus amyloliquefaciens***

Briefly, extracts of *Bacillus amyloliquefaciens* were shown to possess nitroreductase activity. To purify this activity, crude cell extracts were subjected to ammonium sulphate, fractionation and anion exchange chromatography. The purified material was subject to N-terminal amino acid sequence analysis and the information obtained used to cloned the gene via a PCR-based strategy. Following determination of its nucleotide sequence the gene was overexpressed in *E. coli* and the resultant recombinant protein purified and characterised see table 1.

This analysis showed that the enzyme had properties which were distinct from that of *E.coli* NfnB. Thus the protein had a more favourable $K_m$ for CB1954 (1.5-fold lower than the E. coli B NfnB) and furthermore converted CB1954 into the 4HX form alone. It also differed from the *E. coli B* NfnB in that the enzyme showed no activity against the prodrug SN23862.

The isolated enzyme/gene represents a significant improvement over the *E.coli* NfnB enzyme with respect to its activity against the prodrug CB1954 ie., it produces only the 4HX derivative and has an improved $K_m$ for CB1954.

A comparison of the amino acid sequence of the isolated enzyme revealed that it shared a very low level of homology to the rat DTD (c. 25%), but exhibited high homology (70% sequence identity) with the predicted product of a gene that has been discovered in the *Bacillus subtilis* genome sequencing project, designated *ywrO*. On this basis, we have designated the cloned *Bacillus amyloliquefaciens* gene *ywrO,* and its encoded enzyme YwrO.

YwrO BAM is a tetrameric flavoprotein (monomeric molecular mass approximately 22.5 kDa by SDS-PAGE, native molecular mass approximately 90 kDa by gel filtration). Although it shares sequence homology with rat DTD it differs in its enzymic properties in that it can use only NADPH as cofactor ($K_m$ 40 $\mu$M). In common with DTD it can reduce CB1954 but not SN23862, reduction of CB1954 resulting in formation of the 4HX product only ($K_m$ 617 $\mu$M, $k_{cat}$ 8.2). It shows a high affinity for the quinone menadione ($K_m$ 3.4 $\mu$M) and has azoreductase and flavin reductase activity ($K_m$ for FMN 53 $\mu$M, $K_m$ for FAD 209 $\mu$M).

In more detail, N-terminal amino acid sequencing of the purified *Bacillus amyloliquefaciens* nitroreductase enzyme resulted in the following sequence,Met-Lys-Val-Leu-Val-Leu-Ala-Val-His-Pro-Asp-Met-Glu-Asn-Ser-Ala-Val-Asn. When this sequence was used to search available protein databases strong homology was noted with the predicted amino acid sequence of a hypothetical protein, YrkL, identified in the *Bacillus subtilis* genome sequencing project. Significant homology was also evident with two proteins, YabF and YheR, identified during the course of the determination of the *Escherichia coli* genome. These three hypothetical proteins shared weak homology with a number of mammalian quinone reductases and NAD(P)H-oxidoreductases, such as the rat DTD.

In view of this observation, a strategy was formulated whereby sequence homology between the identified bacterial proteins, together with the

determined N-terminal amino acid sequence of the discovered *Bacillus amyloliquefaciens* enzyme, was used to amplify a region of the desired encoding gene from the *Bacillus amyloliquefaciens* genome. The one primer utilised in PCR was a degenerate oligonucleotide sequence which corresponded to a DNA sequence capable of coding for the N-terminal octa-peptide Val-His-Pro-Asp-Met-Glu-Asn. It was composed of the following nucleotides, 5'-GTNCAYCCNGATATGGARAA-3', where Y indicates the presence of a T or C, R indicates the presence of A or G, and N indicates the presence of either T, C, G or A. The second primer was based on the hypothetical sequence His-Gly-Trp-Ala-Tyr-Gly which was found to be entirely conserved between the hypothetical bacterial proteins YrkL (*Bacillus subtilis*) and YabF (*E.coli*), and partially conserved in YheR (*E.coli*). The degenerate oligonucleotide mixture synthesised corresponded to the antisense DNA coding strand, viz., 5'-CCRTANGCCCANCCRTG-3'.

| | | |
|---|---|---|
| E.coli | YheR (90-95) | Arg **Gly** Phe **Ala** Ser **Gly** |
| E.coli | YabF (84-89) | **His Gly Trp Ala Tyr Gly** |
| *B.subtilis* | YrkL (85-90) | **His Gly Trp Ala Tyr Gly** |

The two primers were employed in PCR using chromosomal DNA isolated *from Bacillus amyloliquefaciens* and an amplified DNA fragment of the expected size (approximately 230 bp) obtained. This was cloned into plasmid pCR2.1TOPO (Invitrogen) and its nucleotide sequence determined. Translation of the sequence obtained demonstrated the presence of an open reading frame which encoded a polypeptide which shared 66% sequence similarity with YrkL.

To obtain the entire structural gene, an approach was employed based on inverse PCR. In essence, *B. amyloliquefaciens* DNA was cleaved with the restriction enzyme *StyI* and the fragments generated circularised through their subsequent incubation with DNA ligase. The ligated DNA was then used as the template for a PCR employing two divergent primers based on

the sequenced 220 bp fragment. These were BamNTR11 (5'-GCTTATTGACCGCTGAG-3') and BamNTR14 (5'-GTACAGTGCGCCTCCGC-3'). A 2.9 kb fragment was generated, cloned into pCR2.1TOPO (Invitrogen) and the sequence of the insert determined. This allowed the identification of the nucleotide sequence of the remaining parts of the *B. amyloliquefaciens* gene. Using this information, a contiguous copy of the entire structural gene was amplified from the *B. amyloliquefaciens* chromosome using primers which encompassed the translational start codon (5'-GGTGTGATACATATGAAAGTATTG-3') and resided 3' to the translational stop codon (5'-CGGGGATTCGAATTCTTTCTCAGG-3'). The primer at the 5'-end of the gene was designed such the sequence immediately 5' to the ATG start codon became CAT. This change created an *Nde*I restriction site (CATATG), thereby allowing the cloning of the gene into the equivalent site of the expression vector pMTL1015. This manipulation facilitated the subsequent overexpression of the gene, as insertion of the gene at this point positions the start codon at an optimum distance from the vector borne ribosome binding site.

The strategy employed to clone the BM YwrO gene could be similarly employed to clone further genes encoding novel nitroreductases. This would involve purifying the desired enzyme activity from a cell lysate, and then determining the N-terminal sequence. The data obtained could then be used to design an oligonucleotide primer corresponding to the sense strand of the DNA encoding part or all of the determined amino acid sequence. This primer could then be used, in conjunction with a second primer, to amplify part of the gene encoding the nitroreductase from the chromosome of the bacterial host using PCR. The second primer would correspond to the antisense strand of an internal portion of the targeted gene. Its design would be based on regions of homology which are conserved amongst the type of nitroreductase family that is sought. Thus, in the case of the DTD-like family, the oligonucleotide would, for example be based on the conserved motif His-Gly-Trp-Ala-Tyr-Gly (ie., amino acid

residues 85-90 in the BS YrkL protein). In the case of the NfnB-like family, the oligonucleotdie could be based on the motif Glu-Arg-Tyr-Val-Pro-Val-Met (ie., amino acid residues 170-176 in the BS YodC protein).

Such amplified fragments could then be cloned and sequenced, and new primers designed based on this sequence to isolate the flanking regions of the gene by PCR. Once these have been cloned and sequenced, the entire, contiguous structural gene may be amplified using primers which extend beyond the 5' and 3' end of the translational start and stop codons.

Cloning of genes encoding novel nitroreductases may also be achieved without recourse to N-terminal sequencing of the enzyme, or even its purification. This would involve basing the sequence of both of the oligonucleotides used in the initial PCR reaction on amino acid sequence motifs conserved amongst the two identified nitroreductase families. Thus, in the case of the NfnB-like family, a sense primer (eg., 5'-ATTTCTAAAGAAGAGCTGACGGAA-3') based on the motif Ile-Ser-Lys-Glu-Glu-Leul-Thr-Glu (ie., amino acid residues 13 to 20 of BS YodC) could be employed with the an antisense primer (eg., 5'-CATTACCGGTACATAGCGTTC-3') based on the sequence motif Glu-Arg-Tyr-Val-Pro-Val-Met (ie., amino acid residues 170 to 176). In the case of the DTD-family a sense primer (eg., 5'-CATCCGGATATGGAAAAT-3') based on the motif His-Pro-Asp-Met-Glu-Asn (ie., amino acid residues to 9 to 14 of BM YwrO) could be employed with the an antisense primer (eg., 5'-TCCATATGCCCATCCATA-3') based on the sequence motif Tyr-Gly-Trp-Ala-Tyr-Gly (ie., amino acid residues 85 to 90). Once amplified, the rest of the gene could be isolated using the same procedure as outlined above.

## Example 2

***Bacillus subtilis*** **Nitroreductases**

As indicated above in Example 1, comparative analysis of the *B. subtilis* genome sequence with the amino acid sequence of the isolated *B. amyloliquefaciens* enzyme demonstrated the existence of an enzyme (YwrO) which shared 70% sequence identity. Unexpectedly, *B.subtilis* was found to possess two homologues, *YrkL* and *YdeQ,* which share 54% and 51% sequence homology, respectively, with the *B. amyloliquefaciens* enzyme. All three enzymes share no homology with the *E.coli B* NfnB. They do, however, exhibit weak similarity (c. 25%) to the rat DT-Diaphorase (DTD). Whilst these proteins share a low level of sequence similarity to DTD, and other mammalian equivalents, they are characteristically smaller. This is because of the absence of an extensive internal protein domain at the N-terminus of the protein. Thus, the functional equivalent domain of the rat DTD between amino acid residues 51 to 82, are missing from the BM YwrO protein. In addition, the rat DTD has an extra COOH-terminal domain. These bacterial enzymes are thus distinct from their mammalian equivalents.

A further analysis of the *B.subtilis* genome, demonstrated that two homologues of the *E.coli* NfnB gene were present. Their encoded proteins (Ydgl and YodC) share a barely detectable level of sequence conservation with EC NfnB, of around 20% sequence identity.

*Bacillus subtilis* was thus found to carry at least 5 different enzymes with nitroreductase activity. These are split into two families, thus;-

| | | |
|---|---|---|
| DTD-like | - | 3 members:- YwrO, YrkL, YdeQ |
| NfnB-like | - | 2 members:- Ydgl, YodC |

**Example 3**

**Recombinant Production of Nitroreductases from *Bacillus subtilis***

The DNA encoding all 5 *B.subtilis* nitroreductase enzymes were cloned

from genomic DNA using PCR and the resultant genes, following authentification by nucleotide sequencing, subcloned into a propriety CAMR expression vector (pMTL1015). The expression clones generated have been used to overproduce each of the 5 proteins and the enzymic activity of each assessed in crude lysates. This analysis has demonstrated that whilst the *B.subtilis* YwrO shares similar properties to the *B. amyloliquefaciens* homologue (ie., converts CB1954 to the 4HX derivative alone, but is inactive against SN23862), YrkL and YdeQ have no activity against either of the two prodrugs tested (CB1954 or SN23862) but they may be active against other prodrugs.

Despite the extremely limited sequence similarity to EC NfnB, YdgI and YodC are active against both CB1954 and SN23862. They do, however, produce both the 2HX and 4HX derivatives of CB1954. Their characterisation has shown that they turn over CB1954 at higher rates than EC NfnB (YodC $k_{cat}$ 58, YdgI $k_{cat}$ 30.3 cf 6 for NfnB). Both show a high affinity for menadione and flavins, but they differ in that whereas YdgI uses both NADH and NADPH, YodC shows a preference for the latter. The native molecular mass of YodC (approximately 90kDa) indicates that it is tetrameric (molecular mass estimated from amino acid sequence and by SDS-PAGE being approximately 22 kDa) whereas YdgI appears to be a dimer in the native state (molecular mass by gel filtration approximately 49 kDa).

These finding are further illustrated in Table 2.

**Example 4**

***Bacillus lautus* & *Bacillus pumilis* nitroreductases**

From 103 soil sample isolates tested, two strains (*Bacillus pumilis* CP044 and *Bacillus lautus* CP060) had been previously chosen as possessing extracts which showed the most rapid reduction of both CB1954 and

SN23862. Purification experiments demonstrated that the activity in both extracts was distributed across three distinct peaks. The presence of more than one enzyme activity is consistent with our discovery of multiple forms of proteins in Bacillus able to turnover prodrugs. Eventual purification of the three enzymes of *B. pumilis* CP044 revealed that no one candidate exhibited properties which were an improvement on the *E.coli* NfnB enzyme. In contrast, the proteins in peak 1 and peak 3 of the *B.lautus* CP060 were determined to offer advantage over NfnB.

Thus, whilst the enzyme in peak 1 did not produce the required 4HX derivative of CB1954, it exhibited a 4-fold lower Km with the prodrug SN23862. The enzyme of peak 3 was, however, deemed to be of greatest value as it converted CB1954 solely into the 4HX derivative and had a Km approximately 4-fold lower than NfnB. Furthermore, it also had activity against SN23862. In this respect it shares the properties of both the *Bacillus* DTD-like family (ie., it produces only the 4HX derivative) and the NfnB-like family (ie., it is active against SN23862) - these findings are illustrated in Table 3.

## Example 5

**N-terminal Sequencing of *B. lautus* Nitroreductase**

Electrophoretic separation of the peak 3 demonstrated that 4 protein bands were present which could account for the observed prodrug activity. All four were subjected to N-terminal amino acid sequencing and the activity localised to the fourth protein band from which the nitroreductase may be purified.

## Example 6

**Detection of Nitroreductase Activity in Thermophile Extracts**

As an alternative source novel enzymes, a preliminary screen of CAMRs

thermophile collection was undertaken. Enzymes from this source may have the advantage of greater stability, and therefore longevity of action. Strains were selected on the basis either of sensitivity to CB1954, or those which are resistant but which impart a yellow/golden coloration to agar containing prodrug.

Two of these strains (*B. thermoflavus* and *B. licheniformis*) generated the cytotoxic 4HX form and were selected for further study.

**Example 7**

**Identification Of Further Nitroreductase Enzymes**

Having identified the two families of nitroreductase in *Bacillus*, a search was undertaken of both finished and unfinished genomes for homologues, using YwrO and YodC/NfnB. On the basis of this search homologues of YwrO were identified in the genomes of *Yersinia pestis* and *Porphyromonas gingivalis,* and homologues of NfnB in the genomes of *Pyrococcus furiosus, Haemophilus influenza, Synechocystis* PCC 6803, *Campylobacter jejuni, Archaeglobus, Helicobacter pylori, Heliocbacter fulgidus* and *Thermus aquaticus.*

In addition to the above, two *E.coli* genes were found to be homologues of rat DTD and YwrO, and were designated Yher and YabF. They were discovered to share the characteristic of YwrO in that they lack the internal protein domain found in the rat DTD enzyme and functional mammalian homologues.

(i) *P.gingivalis* YwrO homologue

*P.gingivalis* YwrO homologue is a dimeric flavoprotein with native molecular mass estimated by gel filtration at 40 kDa. Although it shares sequence homology with DTD and forms only the 4HX reduction product of CB1954

($K_m$ 1200$\mu$M, $k_{cat}$ 3.2), it differs from DTD in that it is active with SN23862 and it can only use NADH as cofactor (cf DTD which can use either NADH or NADPH and is inactive with SN23862). It can reduce azodyes but it is inactive with menadione or flavins.

(ii) *C.jejuni* NfnB homologue

*C.jejuni* NfnB homologue produces only the 4HX reduction product of CB1954 ($K_m$ 143 $\mu$M, $k_{cat}$ 11.2) using NADPH as cofactor and it is also active with SN23862. It can use the quinone menadione as substrate as well as azodyes and the flavins FMN and FAD.

(iii) *Archaeoglobus fulgidus* NfnB homologue

*Archaeoglobus fulgidus* NfnB homologue is a dimeric flavoprotein of 42 kDa native molecular mass, producing the 4HX derivative of CB1954 only ($K_m$ 690 $\mu$M, $k_{cat}$ 56.2) using NADPH as cofactor. It is also active with SN23862 and menadione ($K_m$ 9 $\mu$M), but does not decolourise azodyes and has only weak flavin reductase activity.

(iv) *H.influenzae* and *H.pylori* NfnB homologues

Both these enzymes are dimeric flavoproteins and form the 4HX reduction product of CB1954 using NADPH in preference to NADH, but have no activity with azodyes. The former also lacks activity with the quinone menadione and flavins FMN or FAD. Both however have weak activity with SN23862 and may be active with other prodrugs.

(v) *Y.pestis* nfnB homologue and *Synechocystis* YwrO homologue

Both these proteins reduce CB1954 but produce only the relatively non-toxic 2HX derivative using NADPH as cofactor. They do however show

activity with SN23862 and the former can also reduce azodyes.

**Example 8**

**Comparison of Nitroreductase Sequences**

We compared the amino acid sequences of nitoreductases according to the invention with each other and with known rat, human and *E.coli* sequences, and the results are illustrated in Figures 1 and 2. In Figure 1, rat, mouse and two human sequences make up the first four lanes for comparison purposes. It is evident that nitroreductases of the invention are lacking a sequence from positions 51-82 of the rat sequence.

In Figure 2, sequences of nitroreductases of the invention are compared with the known *E.coli* sequence, which is designated nfmB in the second-to-last lane.

The invention thus provides nitroreductase enzymes, DNA and genes therefor and methods of obtaining such enzymes and of using the enzymes and DNA coding therefor in clinical applications.

| ENZYME ACTIVITY | M.Wt (kDa) | CB1954 | | SN23862 |
|---|---|---|---|---|
| | | Product | Km | Km |
| *B. pumilis* CP044 | | | | |
| Peak 1 | ND | 4HX | v. low | ND |
| Peak 2 | ND | 4HX | >1000 | ND |
| Peak 3 | ND | 2/4HX | 999 | ND |
| *B.lautus* CP060 | | | | |
| Peak 1 | 35 | 2HX | 211 | 325 |
| Peak 2 | 42 | 4HX | >2000 | none |
| **Peak 3** | **47** | **4HX** | **257** | **active** |

**Table 3:** Fractionation of nitroreductase activity in cell extracts of *Bacillus lautus* and *Bacillus pumilis*

| STRAIN | CB1954 | | | SN23862 | |
|---|---|---|---|---|---|
| | Product | NADH | NADPH | NADH | NADPH |
| 1078 | 2/4HX | 13.8 | 22.6 | 8.5 | 17.6 |
| **2122[a]** | **2/4HX** | **36.6** | **56.0** | **33.4** | **62.8** |
| **6012 [b]** | **4>2HX** | **15.2** | **37.8** | **8.2** | **35.2** |
| 6013 [c] | 2HX | 9.8 | 49.4 | 6.4 | 39.0 |
| 6031 [d] | 2HX | 11.9 | 42.1 | 8.2 | 33.8 |
| 6036 | 2HX | 10.7 | 26.7 | 7.3 | 26.2 |
| 6044 | 2HX | 4.0 | 21.3 | 4.5 | 9.9 |

**Table 4:** Characteristics of nitroreductase activity of thermophiles identified as being sensitive to CB1954 [Identified as *Bacillus thermoflavus* [a], *Bacillus licheniformis* [b], *Bacillus licheniformis* [c], *Bacillus alkophilus* [d]]

| ENZYME ACTIVITY | M.Wt (kDa) | CB1954 | | SN23862 |
|---|---|---|---|---|
| | | Product | Km | Km |
| *B. pumilis* CP044 | | | | |
| Peak 1 | ND | 4HX | v. low | ND |
| Peak 2 | ND | 4HX | >1000 | ND |
| Peak 3 | ND | 2/4HX | 999 | ND |
| *B.lautus* CP060 | | | | |
| Peak 1 | 35 | 2HX | 211 | 325 |
| Peak 2 | 42 | 4HX | >2000 | none |
| **Peak 3** | **47** | **4HX** | **257** | **active** |

**Table 3:** Fractionation of nitroreductase activity in cell extracts of *Bacillus lautus* and *Bacillus pumilis*

| STRAIN | CB1954 | | | SN23862 | |
|---|---|---|---|---|---|
| | Product | NADH | NADPH | NADH | NADPH |
| 1078 | 2/4HX | 13.8 | 22.6 | 8.5 | 17.6 |
| **2122[a]** | **2/4HX** | **36.6** | **56.0** | **33.4** | **62.8** |
| **6012 [b]** | **4>2HX** | **15.2** | **37.8** | **8.2** | **35.2** |
| 6013 [c] | 2HX | 9.8 | 49.4 | 6.4 | 39.0 |
| 6031 [d] | 2HX | 11.9 | 42.1 | 8.2 | 33.8 |
| 6036 | 2HX | 10.7 | 26.7 | 7.3 | 26.2 |
| 6044 | 2HX | 4.0 | 21.3 | 4.5 | 9.9 |

**Table 4:** Characteristics of nitroreductase activity of thermophiles identified as being sensitive to CB1954 [Identified as *Bacillus thermoflavus* [a], *Bacillus licheniformis* [b], *Bacillus licheniformis* [c], *Bacillus alkophilus* [d]]

## CLAIMS

1. A nitroreductase characterised in that it preferentially reduces CB1954 to a cytotoxic 4-hydroxylamine (4HX) derivative instead of a non-cytotoxic 2-hydroxylamine derivative.

2. A nitroreductase according to Claim 1 further characterised in that it reduces CB1954 to the 4HX derivative with a $K_m$ of less than 700 micromolar.

3. A nitroreductase according to Claim 1 or 2 further characterised in that it is NADPH dependant.

4. A nitroreductase according to any of Claims 1 to 3, further characterised in that it reduces CB1954 to a cytotoxic 4-hydroxylamine (4HX) derivative substantially without producing the non-cytotoxic 2-hydroxylamine derivative.

5. A nitroreductase according to any of Claims 1 to 4 which reduces the prodrug to the toxic derivative with a K*cat* of at least 8.

6. A nitroreductase according to any of Claims 1 to 5, which reduces CB1954 or an analogue thereof to a toxic derivative, shares at least 50% sequence identity with the rat DTD sequence and does not contain a domain that is the same as or corresponds to amino acids 51 to 82 of the rat DTD sequence.

7. A nitroreductase characterised in that it reduces a prodrug to a toxic derivative with a $K_m$ of less 700 micromolar, wherein the prodrug is selected from CB1954 and analogues thereof.

8. A nitroreductase according to Claim 7 which reduces the prodrug to

the toxic derivative with a $K_m$ of less 300 micromolar.

9. A nitroreductase according to Claim 7 or 8 which reduces the prodrug to the toxic derivative with a K*cat* of at least 8.

10. A nitroreductase according to Claim 9 which reduces the prodrug to the toxic derivative with a K*cat* of at least 10.

11. A nitroreductase according to any of Claims 7 to 10, further characterised in that it reduces CB1954 to a toxic derivative, it reduces SN23862 to a toxic derivative, it can use both NADH and NADPH as electron donor and in that it shares no more than 30% sequence identity with the *E.coli* NfnB sequence.

12. A nitroreductase according to any of Claims 7 to 11 further characterised in that it shares at least 50% sequence identity with the rat DTD sequence and in that it does not contain a domain that is the same as or corresponds to amino acids 51 to 82 of the rat DTD sequence.

13. A nitroreductase characterised in that it reduces a prodrug to a toxic derivative with a K*cat* of at least 8.

14. A nitroreductase according to Claim 13, further characterised in that it reduces CB1954 to a toxic derivative, it reduces SN23862 to a toxic derivative, it can use both NADH and NADPH as electron donor and in that it shares no more than 30% sequence identity with the *E.coli* NfnB sequence.

15. A nitroreductase according to Claim 13 or 14, further characterised in that it reduces CB1954 or an analogue thereof to a toxic derivative, in that it shares at least 50% sequence identity with the rat DTD sequence and in that it does not contain a domain that is the same as or corresponds

to amino acids 51 to 82 of the rat DTD sequence.

16. A nitroreductase characterised in that it reduces CB1954 to a toxic derivative, it reduces SN23862 to a toxic derivative, it can use both NADH and NADPH as electron donor and in that it shares no more than 30% sequence identity with the *E.coli* NfnB sequence.

17. A nitroreductase according to Claim 16, wherein the sequence identity is about 25% or less.

18. A nitroreductase characterised in that it reduces CB1954 or an analogue thereof to a toxic derivative, in that it shares at least 50% sequence identity with the rat DTD sequence and in that it does not contain a domain that is the same as or corresponds to amino acids 51 to 82 of the rat DTD sequence.

19. Use of a DNA sequence coding for a nitroreductase according to any preceding Claim in manufacture of a medicament for prodrug therapy.

20. A viral vector, comprising

(a) a DNA encoding nitroreductase according to any of Claims 1 to 18 operatively coupled to

(b) a promoter for expression of the DNA.

21. A mini-gene comprising

(a) a DNA encoding nitroreductase according to any of Claims 1 to 18 operatively coupled to

(b) a promoter for expression of the DNA.

22. A pharmaceutical composition comprising a nitroreductase according to any of Claims 1 to 18 in combination with a pharmaceutically acceptable carrier.

23. A pharmaceutical composition for use in a directed-enzyme prodrug therapy, comprising a pharmaceutically acceptable carrier and a compound for converting a prodrug into a drug, wherein a compound comprises a nitroreductase according to any of Claims 1 to 18 conjugated to a targeting moiety.

24. A pharmaceutical composition according to Claim 23 wherein the targeting moiety comprises an antibody specific for a target cell.

25. A pharmaceutical composition according to Claim 23 wherein the targeting moiety is a moiety preferentially accumulated by or taking up by a target cell.

26. A method of preparing a nitroreductase, comprising expressing a gene in a bacterial cell, wherein the gene codes for a nitroreductase according to any of Claims 1 to 18.

27. Use of a nitroreductase according to any of Claims 1-18 in manufacture of a medicament for anti-tumour therapy.

28. Use of a compound comprising a nitroreductase according to any of Claims 1 to 18 conjugated to a targeting moiety in manufacture of a medicament for anti-tumour therapy.

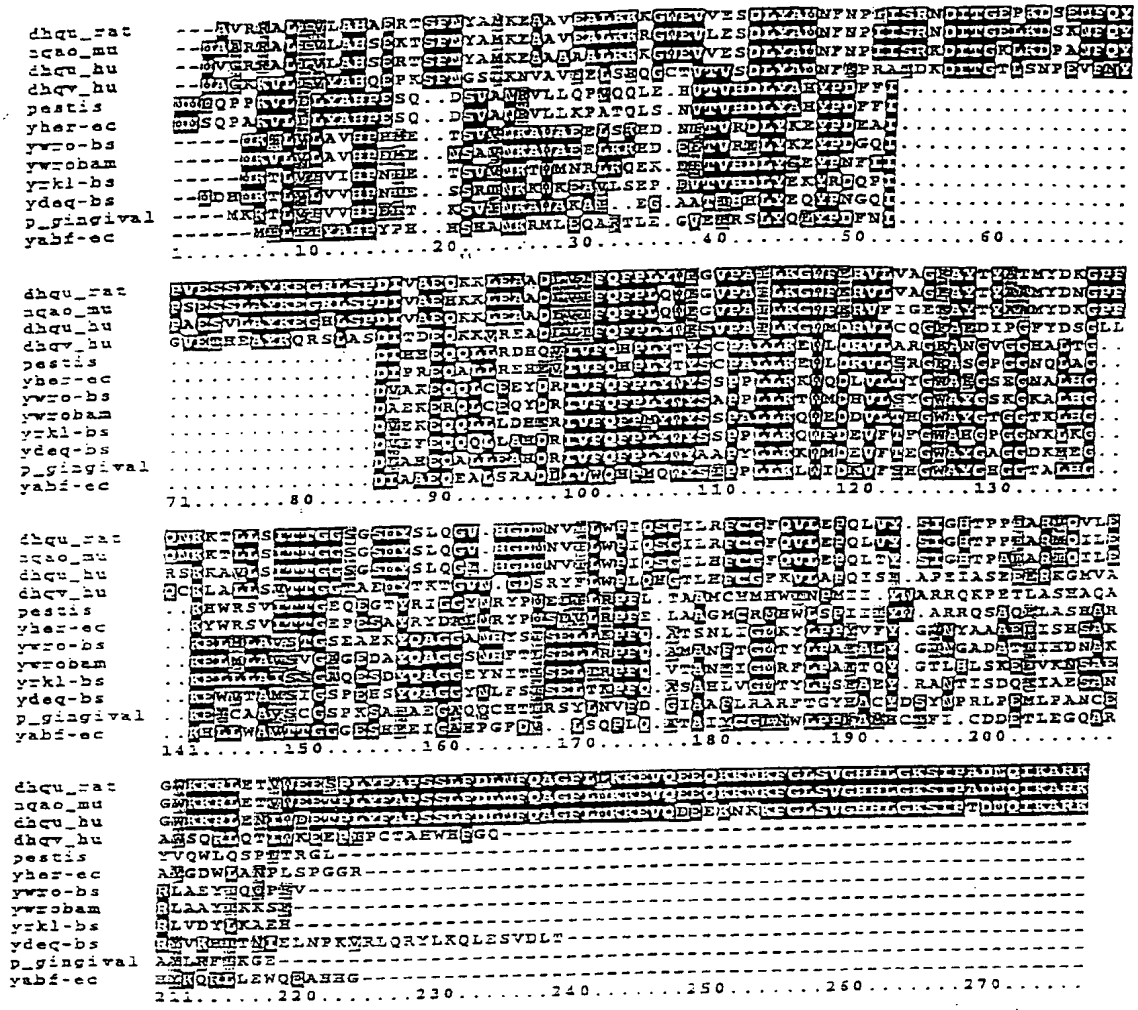
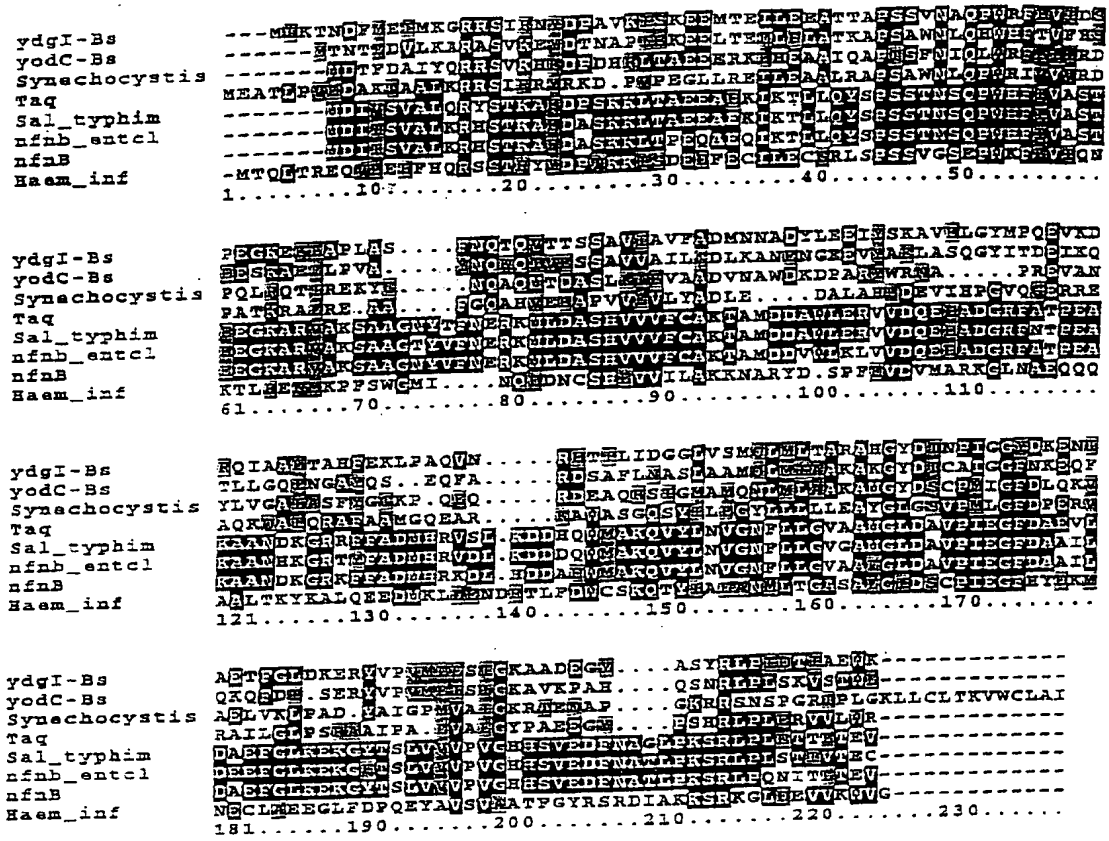


Fig. 1

DTD-Like Proteins

The aligned proteins are: NQO1_rat, NAD(P)H-quinone oxidoreductase 1 (brown rat); NQO1_mouse, NAD(P)H-quinone oxidoreductase 1 (mouse); NQO1_human, NAD(P)H-quinone oxidoreductase 1 (human); Yersinia, NQO2_human, NAD(P)H-quinone oxidoreductase 2 (human); yheR_Ecoli, yheR (*Escherichia coli*); ywrOsubtil, ywrO un-named homologue (*Yersinia pestis*); YrkLsubtil, yrkl (*Bacillus subtilis*); ywrO_amylo, ywrO (*Bacillus amyloliquefaciens*); ydeQsubtil, ydeQ (Bacillus subtilis); Porph_ging, un-named homologue (*Porphyromonas gingivalis*), and; yabF_Ecoli, yabF (*Escherichia coli*)




NfnB-Like Proteins

The aligned proteins are: ydgI-BS, ydgI (*Bacillus subtilis*); yodC-Bs, yodC (*Bacillus subtilis*); Synechocystis, drgA (*Synechocystis* PCC 6803); Taq, NOX_THETH (*Thermus aquaticus*); Sal_typhim, nfnB (*Salmonella typhimurium*); nfnb_entcl, nfnB (*Enterobacter cloacae*); nfnB, nfnB (*Escherichia coli* B), and; Haem_inf, YC78_HAEIN (*Haemophilus influenzae*).

SEQUENCE LISTING

<110> Microbiological Research Authority

<120> Nitroreductase Enzymes

<130> gws/21226-seq

<140>
<141>

<160> 27

<170> PatentIn Ver. 2.1

<210> 1
<211> 525
<212> DNA
<213> Bacillus amyloliquefaciens

<220>
<221> CDS
<222> (1)..(525)

<400> 1

```
gtg aaa gta ttg gta tta gcg gtt cac cct gac atg gag aac tca gcg   48
Met Lys Val Leu Val Leu Ala Val His Pro Asp Met Glu Asn Ser Ala
  1               5                  10                  15

gtc aat aag gca tgg gca gaa gaa tta aaa aaa cat gat gaa ctc acg   96
Val Asn Lys Ala Trp Ala Glu Glu Leu Lys Lys His Asp Glu Leu Thr
            20                  25                  30

gtc cgt gag ctt tat aaa gaa tat ccg gac ggg caa atc gat gcg gaa  144
Val Arg Glu Leu Tyr Lys Glu Tyr Pro Asp Gly Gln Ile Asp Ala Glu
        35                  40                  45

aag gaa cgt cag ctg tgt gaa cag tat gac cgg atc gta ttt caa ttt  192
Lys Glu Arg Gln Leu Cys Glu Gln Tyr Asp Arg Ile Val Phe Gln Phe
    50                  55                  60

ccg ctg tat tgg tac agt gcg cct ccg ctt tta aaa aca tgg atg gat  240
Pro Leu Tyr Trp Tyr Ser Ala Pro Pro Leu Leu Lys Thr Trp Met Asp
65                  70                  75                  80

cat gtg ctg tcg tac ggc tgg gcc tac ggc tcc aaa gga aag gcg ctg  288
His Val Leu Ser Tyr Gly Trp Ala Tyr Gly Ser Lys Gly Lys Ala Leu
                85                  90                  95

cat ggc aaa gaa ttg atg ctg gct gtt tcc gta ggt gcc gga gag gat  336
His Gly Lys Glu Leu Met Leu Ala Val Ser Val Gly Ala Gly Glu Asp
            100                 105                 110

gca tac cag gca gga ggg tca aac cac ttt aca ttg agc gag ctg tta  384
Ala Tyr Gln Ala Gly Gly Ser Asn His Phe Thr Leu Ser Glu Leu Leu
        115                 120                 125

agg ccg ttt cag gca atg gct aat ttt aca ggt atg acc tat ttg ccg  432
Arg Pro Phe Gln Ala Met Ala Asn Phe Thr Gly Met Thr Tyr Leu Pro
    130                 135                 140

gct ttc gcg ctg tac ggt gta aat ggg gcg gat gcg acg gat att cat  480
Ala Phe Ala Leu Tyr Gly Val Asn Gly Ala Asp Ala Thr Asp Ile His
145                 150                 155                 160

gac aat gcc aaa cgt ctg gct gct tac ata aag aaa tca ttt taa      525
```

```
Asp Asn Ala Lys Arg Leu Ala Ala Tyr Ile Lys Lys Ser Phe
                165                 170                 175


<210> 2
<211> 175
<212> PRT
<213> Bacillus amyloliquefaciens

<400> 2
Val Lys Val Leu Val Leu Ala Val His Pro Asp Met Glu Asn Ser Ala
  1               5                  10                  15

Val Asn Lys Ala Trp Ala Glu Glu Leu Lys Lys His Asp Glu Leu Thr
            20                  25                  30

Val Arg Glu Leu Tyr Lys Glu Tyr Pro Asp Gly Gln Ile Asp Ala Glu
        35                  40                  45

Lys Glu Arg Gln Leu Cys Glu Gln Tyr Asp Arg Ile Val Phe Gln Phe
    50                  55                  60

Pro Leu Tyr Trp Tyr Ser Ala Pro Pro Leu Leu Lys Thr Trp Met Asp
65                  70                  75                  80

His Val Leu Ser Tyr Gly Trp Ala Tyr Gly Ser Lys Gly Lys Ala Leu
                85                  90                  95

His Gly Lys Glu Leu Met Leu Ala Val Ser Val Gly Ala Gly Glu Asp
            100                 105                 110

Ala Tyr Gln Ala Gly Gly Ser Asn His Phe Thr Leu Ser Glu Leu Leu
        115                 120                 125

Arg Pro Phe Gln Ala Met Ala Asn Phe Thr Gly Met Thr Tyr Leu Pro
    130                 135                 140

Ala Phe Ala Leu Tyr Gly Val Asn Gly Ala Asp Ala Thr Asp Ile His
145                 150                 155                 160

Asp Asn Ala Lys Arg Leu Ala Ala Tyr Ile Lys Lys Ser Phe
                165                 170                 175


<210> 3
<211> 528
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(528)

<400> 3
atg aaa ata ttg gtt ttg gca gtg cat cct cat atg gag acc tca gtt   48
Met Lys Ile Leu Val Leu Ala Val His Pro His Met Glu Thr Ser Val
  1               5                  10                  15

gtt aat aag gcg tgg gct gag gaa ttg agt aaa cat gac aat atc aca   96
Val Asn Lys Ala Trp Ala Glu Glu Leu Ser Lys His Asp Asn Ile Thr
            20                  25                  30

gta cgg gat ctt tat aag gaa tac ccg gat gaa gcg ata gat gtt gcg   144
Val Arg Asp Leu Tyr Lys Glu Tyr Pro Asp Glu Ala Ile Asp Val Ala
        35                  40                  45

aag gaa cag cag ctg tgc gag gaa tac gat cgg att gtc ttt caa ttc   192
```

```
Lys Glu Gln Gln Leu Cys Glu Glu Tyr Asp Arg Ile Val Phe Gln Phe
     50                  55                  60

ccg cta tat tgg tac agc tct ccg ccg ctc ttg aaa aaa tgg cag gat   240
Pro Leu Tyr Trp Tyr Ser Ser Pro Pro Leu Leu Lys Lys Trp Gln Asp
 65                  70                  75                  80

ctt gtg ctg act tat ggc tgg gct ttt ggt tca gaa gga aat gcc ttg   288
Leu Val Leu Thr Tyr Gly Trp Ala Phe Gly Ser Glu Gly Asn Ala Leu
                85                  90                  95

cat ggc aag gag ctg atg ctg gct gta tca aca ggg agc gaa gca gaa   336
His Gly Lys Glu Leu Met Leu Ala Val Ser Thr Gly Ser Glu Ala Glu
            100                 105                 110

aaa tat caa gcg ggc gga gca aat cat tac tcg atc agt gag cta ttg   384
Lys Tyr Gln Ala Gly Gly Ala Asn His Tyr Ser Ile Ser Glu Leu Leu
        115                 120                 125

aaa cca ttt cag gcc acg agt aat ctg atc ggc atg aag tat ctg cct   432
Lys Pro Phe Gln Ala Thr Ser Asn Leu Ile Gly Met Lys Tyr Leu Pro
    130                 135                 140

cca tat gtg ttc tat ggc gtg aat tat gca gct gca gag gat att tct   480
Pro Tyr Val Phe Tyr Gly Val Asn Tyr Ala Ala Ala Glu Asp Ile Ser
145                 150                 155                 160

cac agt gca aaa cgg tta gcc gaa tac atc cag cag cct ttt gtt taa   528
His Ser Ala Lys Arg Leu Ala Glu Tyr Ile Gln Gln Pro Phe Val
                165                 170                 175
```

<210> 4
<211> 176
<212> PRT
<213> Bacillus subtilis

```
<400> 4
Met Lys Ile Leu Val Leu Ala Val His Pro His Met Glu Thr Ser Val
  1               5                  10                  15

Val Asn Lys Ala Trp Ala Glu Glu Leu Ser Lys His Asp Asn Ile Thr
            20                  25                  30

Val Arg Asp Leu Tyr Lys Glu Tyr Pro Asp Glu Ala Ile Asp Val Ala
        35                  40                  45

Lys Glu Gln Gln Leu Cys Glu Glu Tyr Asp Arg Ile Val Phe Gln Phe
    50                  55                  60

Pro Leu Tyr Trp Tyr Ser Ser Pro Pro Leu Leu Lys Lys Trp Gln Asp
65                  70                  75                  80

Leu Val Leu Thr Tyr Gly Trp Ala Phe Gly Ser Glu Gly Asn Ala Leu
                85                  90                  95

His Gly Lys Glu Leu Met Leu Ala Val Ser Thr Gly Ser Glu Ala Glu
            100                 105                 110

Lys Tyr Gln Ala Gly Gly Ala Asn His Tyr Ser Ile Ser Glu Leu Leu
        115                 120                 125

Lys Pro Phe Gln Ala Thr Ser Asn Leu Ile Gly Met Lys Tyr Leu Pro
    130                 135                 140

Pro Tyr Val Phe Tyr Gly Val Asn Tyr Ala Ala Ala Glu Asp Ile Ser
145                 150                 155                 160
```

```
His Ser Ala Lys Arg Leu Ala Glu Tyr Ile Gln Gln Pro Phe Val
                165                 170                 175


<210> 5
<211> 525
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(525)

<400> 5
atg aaa aca tta gtt atc gtt ata cat cct aat ttg gaa acg tct gtt   48
Met Lys Thr Leu Val Ile Val Ile His Pro Asn Leu Glu Thr Ser Val
  1               5                  10                  15

gtc aac aaa acc tgg atg aat cgt tta aag caa gag aaa gac att acg   96
Val Asn Lys Thr Trp Met Asn Arg Leu Lys Gln Glu Lys Asp Ile Thr
            20                  25                  30

gtt cat gac ctg tac ggt gaa tac cct aat ttt atc att gat gta gaa   144
Val His Asp Leu Tyr Gly Glu Tyr Pro Asn Phe Ile Ile Asp Val Glu
        35                  40                  45

aaa gag cag cag ctc ctg tta gat cat gag cgt atc gtt ttt cag ttc   192
Lys Glu Gln Gln Leu Leu Leu Asp His Glu Arg Ile Val Phe Gln Phe
    50                  55                  60

cca atg tat tgg tac agc agt ccc gcg tta ctc aaa caa tgg gaa gat   240
Pro Met Tyr Trp Tyr Ser Ser Pro Ala Leu Leu Lys Gln Trp Glu Asp
65                  70                  75                  80

gat gtg tta aca cat ggc tgg gct tat gga act gga gga act aaa ttg   288
Asp Val Leu Thr His Gly Trp Ala Tyr Gly Thr Gly Gly Thr Lys Leu
                85                  90                  95

cat gga aaa gaa cta ctc tta gct atc tcc tca ggc gca cag gaa tct   336
His Gly Lys Glu Leu Leu Leu Ala Ile Ser Ser Gly Ala Gln Glu Ser
            100                 105                 110

gat tat caa gca ggc gga gaa tat aat atc acg atc agc gag ctt atc   384
Asp Tyr Gln Ala Gly Gly Glu Tyr Asn Ile Thr Ile Ser Glu Leu Ile
        115                 120                 125

aga ccg ttt caa gtc act gct aac tat ata gga atg cgt ttt ctt cct   432
Arg Pro Phe Gln Val Thr Ala Asn Tyr Ile Gly Met Arg Phe Leu Pro
    130                 135                 140

gcg ttt aca caa tat ggg aca ctt cat ctt tca aaa gaa gat gtt aag   480
Ala Phe Thr Gln Tyr Gly Thr Leu His Leu Ser Lys Glu Asp Val Lys
145                 150                 155                 160

aac agt gcg gag aga ttg gtt gac tat ctt aaa gcc gag cat taa       525
Asn Ser Ala Glu Arg Leu Val Asp Tyr Leu Lys Ala Glu His
                165                 170                 175


<210> 6
<211> 175
<212> PRT
<213> Bacillus subtilis

<400> 6
Met Lys Thr Leu Val Ile Val Ile His Pro Asn Leu Glu Thr Ser Val
  1               5                  10                  15
```

```
Val Asn Lys Thr Trp Met Asn Arg Leu Lys Gln Glu Lys Asp Ile Thr
                20                  25                  30

Val His Asp Leu Tyr Gly Glu Tyr Pro Asn Phe Ile Ile Asp Val Glu
        35                  40                  45

Lys Glu Gln Gln Leu Leu Leu Asp His Glu Arg Ile Val Phe Gln Phe
    50                  55                  60

Pro Met Tyr Trp Tyr Ser Ser Pro Ala Leu Leu Lys Gln Trp Glu Asp
65                  70                  75                  80

Asp Val Leu Thr His Gly Trp Ala Tyr Gly Thr Gly Gly Thr Lys Leu
                85                  90                  95

His Gly Lys Glu Leu Leu Leu Ala Ile Ser Ser Gly Ala Gln Glu Ser
            100                 105                 110

Asp Tyr Gln Ala Gly Gly Glu Tyr Asn Ile Thr Ile Ser Glu Leu Ile
        115                 120                 125

Arg Pro Phe Gln Val Thr Ala Asn Tyr Ile Gly Met Arg Phe Leu Pro
    130                 135                 140

Ala Phe Thr Gln Tyr Gly Thr Leu His Leu Ser Lys Glu Asp Val Lys
145                 150                 155                 160

Asn Ser Ala Glu Arg Leu Val Asp Tyr Leu Lys Ala Glu His
                165                 170                 175
```

<210> 7
<211> 594
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(594)

<400> 7

```
atg gat cat atg aaa aca ctc gta ctc gtt gta cat ccg aat ata gaa   48
Met Asp His Met Lys Thr Leu Val Leu Val Val His Pro Asn Ile Glu
  1               5                  10                  15

tcc tct cgt atc aat aaa aag tgg aaa gaa gcc gtt tta agt gaa cca   96
Ser Ser Arg Ile Asn Lys Lys Trp Lys Glu Ala Val Leu Ser Glu Pro
            20                  25                  30

gat gta act gtc cat gat ctt tat gaa aaa tat cgc gat caa cca att   144
Asp Val Thr Val His Asp Leu Tyr Glu Lys Tyr Arg Asp Gln Pro Ile
        35                  40                  45

gat gtg gaa ttt gaa caa cag cag ctc ctg gcc cat gac cgt atc gtt   192
Asp Val Glu Phe Glu Gln Gln Gln Leu Leu Ala His Asp Arg Ile Val
    50                  55                  60

ttt cag ttt cca tta tac tgg tac agc agc cca ccg ctt tta aaa cag   240
Phe Gln Phe Pro Leu Tyr Trp Tyr Ser Ser Pro Pro Leu Leu Lys Gln
65                  70                  75                  80

tgg ttt gat gaa gtg ttt acg ttt ggc tgg gct cat ggt ccc ggc gga   288
Trp Phe Asp Glu Val Phe Thr Phe Gly Trp Ala His Gly Pro Gly Gly
                85                  90                  95

aat aaa ttg aag ggg aaa gag tgg gta act gcc atg tcc atc ggt tca   336
Asn Lys Leu Lys Gly Lys Glu Trp Val Thr Ala Met Ser Ile Gly Ser
```

```
            100                 105                 110

cct gaa cac tct tat caa gcc ggc gga tat aac ttg ttt tcg ata agc   384
Pro Glu His Ser Tyr Gln Ala Gly Gly Tyr Asn Leu Phe Ser Ile Ser
        115                 120                 125

gag ctg aca aaa ccg ttc caa gca tct gcc cat tta gta ggc atg acc   432
Glu Leu Thr Lys Pro Phe Gln Ala Ser Ala His Leu Val Gly Met Thr
    130                 135                 140

tat ctg cct tcc ttt gcc gaa tat cgc gcc aat aca atc agt gac caa   480
Tyr Leu Pro Ser Phe Ala Glu Tyr Arg Ala Asn Thr Ile Ser Asp Gln
145                 150                 155                 160

gaa att gcc gaa agt gcg aat cgg tat gta aag cat att aca aat ata   528
Glu Ile Ala Glu Ser Ala Asn Arg Tyr Val Lys His Ile Thr Asn Ile
                165                 170                 175

gaa tta aac ccg aag gtt cgc ctg caa agg tat ttg aaa cag ctg gag   576
Glu Leu Asn Pro Lys Val Arg Leu Gln Arg Tyr Leu Lys Gln Leu Glu
            180                 185                 190

agt gtc gat tta aca taa                                          594
Ser Val Asp Leu Thr
        195
```

<210> 8
<211> 198
<212> PRT
<213> Bacillus subtilis

<400> 8

```
Met Asp His Met Lys Thr Leu Val Leu Val Val His Pro Asn Ile Glu
  1               5                  10                  15

Ser Ser Arg Ile Asn Lys Lys Trp Lys Glu Ala Val Leu Ser Glu Pro
            20                  25                  30

Asp Val Thr Val His Asp Leu Tyr Glu Lys Tyr Arg Asp Gln Pro Ile
        35                  40                  45

Asp Val Glu Phe Glu Gln Gln Gln Leu Leu Ala His Asp Arg Ile Val
    50                  55                  60

Phe Gln Phe Pro Leu Tyr Trp Tyr Ser Ser Pro Pro Leu Leu Lys Gln
65                  70                  75                  80

Trp Phe Asp Glu Val Phe Thr Phe Gly Trp Ala His Gly Pro Gly Gly
                85                  90                  95

Asn Lys Leu Lys Gly Lys Glu Trp Val Thr Ala Met Ser Ile Gly Ser
            100                 105                 110

Pro Glu His Ser Tyr Gln Ala Gly Gly Tyr Asn Leu Phe Ser Ile Ser
        115                 120                 125

Glu Leu Thr Lys Pro Phe Gln Ala Ser Ala His Leu Val Gly Met Thr
    130                 135                 140

Tyr Leu Pro Ser Phe Ala Glu Tyr Arg Ala Asn Thr Ile Ser Asp Gln
145                 150                 155                 160

Glu Ile Ala Glu Ser Ala Asn Arg Tyr Val Lys His Ile Thr Asn Ile
                165                 170                 175

Glu Leu Asn Pro Lys Val Arg Leu Gln Arg Tyr Leu Lys Gln Leu Glu
```

```
            180                 185                 190

Ser Val Asp Leu Thr
        195


<210> 9
<211> 630
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(630)

<400> 9
atg atc aaa aca aac gat ttt atg gaa att atg aaa ggc cgc cgt tct   48
Met Ile Lys Thr Asn Asp Phe Met Glu Ile Met Lys Gly Arg Arg Ser
  1               5                  10                  15

atc cgc aac tat gat ccg gca gta aaa atc agc aaa gaa gaa atg aca   96
Ile Arg Asn Tyr Asp Pro Ala Val Lys Ile Ser Lys Glu Glu Met Thr
             20                  25                  30

gag atc tta gag gaa gca aca act gcc cca tct tct gtt aac gcg cag   144
Glu Ile Leu Glu Glu Ala Thr Thr Ala Pro Ser Ser Val Asn Ala Gln
         35                  40                  45

cca tgg cgt ttt ctt gtc att gac agc ccg gaa gga aaa gaa aag ctc   192
Pro Trp Arg Phe Leu Val Ile Asp Ser Pro Glu Gly Lys Glu Lys Leu
     50                  55                  60

gca ccg ctt gca agc ttt aac caa aca caa gtc aca aca tca tct gct   240
Ala Pro Leu Ala Ser Phe Asn Gln Thr Gln Val Thr Thr Ser Ser Ala
 65                  70                  75                  80

gtc atc gct gta ttt gca gac atg aac aac gca gac tat cta gaa gaa   288
Val Ile Ala Val Phe Ala Asp Met Asn Asn Ala Asp Tyr Leu Glu Glu
                 85                  90                  95

atc tat tca aaa gcc gtg gaa ctt ggt tac atg ccg cag gag gtc aaa   336
Ile Tyr Ser Lys Ala Val Glu Leu Gly Tyr Met Pro Gln Glu Val Lys
            100                 105                 110

gac aga caa atc gcc gcg ctg acc gca cat ttt gaa aag ctt ccg gca   384
Asp Arg Gln Ile Ala Ala Leu Thr Ala His Phe Glu Lys Leu Pro Ala
        115                 120                 125

cag gtc aac cgt gaa acg atc ctg att gac gga ggt ctt gtt tcc atg   432
Gln Val Asn Arg Glu Thr Ile Leu Ile Asp Gly Gly Leu Val Ser Met
    130                 135                 140

cag ctg atg ctg act gca cgc gcg cat ggc tac gat aca aac ccg atc   480
Gln Leu Met Leu Thr Ala Arg Ala His Gly Tyr Asp Thr Asn Pro Ile
145                 150                 155                 160

ggc gga tac gat aaa gaa aac atc gcg gaa acc ttc gga tta gat aaa   528
Gly Gly Tyr Asp Lys Glu Asn Ile Ala Glu Thr Phe Gly Leu Asp Lys
                165                 170                 175

gaa cgt tat gta ccg gtt atg cta ctt tct atc gga aaa gca gca gac   576
Glu Arg Tyr Val Pro Val Met Leu Leu Ser Ile Gly Lys Ala Ala Asp
            180                 185                 190

gaa ggc tat gct tcc tac cgt ctg ccg att gat aca att gca gaa tgg   624
Glu Gly Tyr Ala Ser Tyr Arg Leu Pro Ile Asp Thr Ile Ala Glu Trp
        195                 200                 205
```

```
aaa taa                                                                  630
Lys
    210


<210> 10
<211> 210
<212> PRT
<213> Bacillus subtilis

<400> 10
Met Ile Lys Thr Asn Asp Phe Met Glu Ile Met Lys Gly Arg Arg Ser
  1               5                  10                  15

Ile Arg Asn Tyr Asp Pro Ala Val Lys Ile Ser Lys Glu Glu Met Thr
            20                  25                  30

Glu Ile Leu Glu Glu Ala Thr Thr Ala Pro Ser Ser Val Asn Ala Gln
        35                  40                  45

Pro Trp Arg Phe Leu Val Ile Asp Ser Pro Glu Gly Lys Glu Lys Leu
    50                  55                  60

Ala Pro Leu Ala Ser Phe Asn Gln Thr Gln Val Thr Thr Ser Ser Ala
65                  70                  75                  80

Val Ile Ala Val Phe Ala Asp Met Asn Asn Ala Asp Tyr Leu Glu Glu
                85                  90                  95

Ile Tyr Ser Lys Ala Val Glu Leu Gly Tyr Met Pro Gln Glu Val Lys
            100                 105                 110

Asp Arg Gln Ile Ala Ala Leu Thr Ala His Phe Glu Lys Leu Pro Ala
        115                 120                 125

Gln Val Asn Arg Glu Thr Ile Leu Ile Asp Gly Gly Leu Val Ser Met
    130                 135                 140

Gln Leu Met Leu Thr Ala Arg Ala His Gly Tyr Asp Thr Asn Pro Ile
145                 150                 155                 160

Gly Gly Tyr Asp Lys Glu Asn Ile Ala Glu Thr Phe Gly Leu Asp Lys
                165                 170                 175

Glu Arg Tyr Val Pro Val Met Leu Leu Ser Ile Gly Lys Ala Ala Asp
            180                 185                 190

Glu Gly Tyr Ala Ser Tyr Arg Leu Pro Ile Asp Thr Ile Ala Glu Trp
        195                 200                 205

Lys
    210


<210> 11
<211> 609
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(609)

<400> 11
atg acg aat act ctg gat gtt tta aaa gca cgt gca tct gta aag gaa    48
Met Thr Asn Thr Leu Asp Val Leu Lys Ala Arg Ala Ser Val Lys Glu
  1               5                  10                  15
```

```
tat gat aca aat gcc ccg atc tct aag gag gag ctg act gag cta tta   96
Tyr Asp Thr Asn Ala Pro Ile Ser Lys Glu Glu Leu Thr Glu Leu Leu
                20                  25                  30

gac ctt gcc act aaa gcg cct tct gct tgg aac ctt cag cat tgg cat  144
Asp Leu Ala Thr Lys Ala Pro Ser Ala Trp Asn Leu Gln His Trp His
            35                  40                  45

ttt aca gta ttc cac agc gat gaa tca aaa gcg gag ctt ctt cct gta  192
Phe Thr Val Phe His Ser Asp Glu Ser Lys Ala Glu Leu Leu Pro Val
        50                  55                  60

gcg tat aat caa aaa caa atc gtt gag tct tct gct gtt gtt gcc att  240
Ala Tyr Asn Gln Lys Gln Ile Val Glu Ser Ser Ala Val Val Ala Ile
    65                  70                  75                  80

tta ggc gat tta aag gca aat gaa aac ggt gaa gaa gtt tat gct gaa  288
Leu Gly Asp Leu Lys Ala Asn Glu Asn Gly Glu Glu Val Tyr Ala Glu
                85                  90                  95

tta gca agc caa ggc tat att acg gat gaa atc aaa caa aca ttg ctc  336
Leu Ala Ser Gln Gly Tyr Ile Thr Asp Glu Ile Lys Gln Thr Leu Leu
            100                 105                 110

ggc caa atc aac ggt gct tac caa agc gag caa ttc gca cgt gat tcc  384
Gly Gln Ile Asn Gly Ala Tyr Gln Ser Glu Gln Phe Ala Arg Asp Ser
        115                 120                 125

gct ttc tta aat gct tct tta gct gct atg cag ctt atg att gcc gca  432
Ala Phe Leu Asn Ala Ser Leu Ala Ala Met Gln Leu Met Ile Ala Ala
    130                 135                 140

aaa gca aaa ggt tat gac act tgc gca atc ggc gga ttt aac aaa gag  480
Lys Ala Lys Gly Tyr Asp Thr Cys Ala Ile Gly Gly Phe Asn Lys Glu
145                 150                 155                 160

cag ttc caa aag caa ttt gat atc agt gag cgc tat gtt ccg gtt atg  528
Gln Phe Gln Lys Gln Phe Asp Ile Ser Glu Arg Tyr Val Pro Val Met
                165                 170                 175

ctt att tca atc ggc aaa gca gtg aag cct gcg cat caa agc aac cgt  576
Leu Ile Ser Ile Gly Lys Ala Val Lys Pro Ala His Gln Ser Asn Arg
            180                 185                 190

ctg ccg ctt tca aaa gta tca act tgg ctg taa                     609
Leu Pro Leu Ser Lys Val Ser Thr Trp Leu
        195                 200
```

<210> 12
<211> 203
<212> PRT
<213> Bacillus subtilis

```
<400> 12
Met Thr Asn Thr Leu Asp Val Leu Lys Ala Arg Ala Ser Val Lys Glu
  1               5                  10                  15

Tyr Asp Thr Asn Ala Pro Ile Ser Lys Glu Glu Leu Thr Glu Leu Leu
            20                  25                  30

Asp Leu Ala Thr Lys Ala Pro Ser Ala Trp Asn Leu Gln His Trp His
        35                  40                  45

Phe Thr Val Phe His Ser Asp Glu Ser Lys Ala Glu Leu Leu Pro Val
    50                  55                  60
```

```
Ala Tyr Asn Gln Lys Gln Ile Val Glu Ser Ser Ala Val Val Ala Ile
 65                  70                  75                  80

Leu Gly Asp Leu Lys Ala Asn Glu Asn Gly Glu Glu Val Tyr Ala Glu
                85                  90                  95

Leu Ala Ser Gln Gly Tyr Ile Thr Asp Glu Ile Lys Gln Thr Leu Leu
            100                 105                 110

Gly Gln Ile Asn Gly Ala Tyr Gln Ser Glu Gln Phe Ala Arg Asp Ser
        115                 120                 125

Ala Phe Leu Asn Ala Ser Leu Ala Ala Met Gln Leu Met Ile Ala Ala
    130                 135                 140

Lys Ala Lys Gly Tyr Asp Thr Cys Ala Ile Gly Gly Phe Asn Lys Glu
145                 150                 155                 160

Gln Phe Gln Lys Gln Phe Asp Ile Ser Glu Arg Tyr Val Pro Val Met
                165                 170                 175

Leu Ile Ser Ile Gly Lys Ala Val Lys Pro Ala His Gln Ser Asn Arg
            180                 185                 190

Leu Pro Leu Ser Lys Val Ser Thr Trp Leu
        195                 200
```

<210> 13
<211> 555
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(555)

<400> 13

```
atg atg tct cag cca gcg aaa gtt ttg ctg ctg tat gcc cat ccg gaa   48
Met Met Ser Gln Pro Ala Lys Val Leu Leu Leu Tyr Ala His Pro Glu
  1               5                  10                  15

tct cag gac tcg gtg gca aac cgg gta ctg ctt aaa ccg gcc acg cag   96
Ser Gln Asp Ser Val Ala Asn Arg Val Leu Leu Lys Pro Ala Thr Gln
             20                  25                  30

ctc agc aat gtt acc gtg cac gac ctt tac gcg cac tat ccc gat ttt  144
Leu Ser Asn Val Thr Val His Asp Leu Tyr Ala His Tyr Pro Asp Phe
         35                  40                  45

ttt att gat atc ccc cgt gag cag gca tta ctg cgc gag cac gag gtg  192
Phe Ile Asp Ile Pro Arg Glu Gln Ala Leu Leu Arg Glu His Glu Val
     50                  55                  60

att gtc ttt cag cat cct ctt tat acc tat agc tgc ccg gcg cta ctg  240
Ile Val Phe Gln His Pro Leu Tyr Thr Tyr Ser Cys Pro Ala Leu Leu
 65                  70                  75                  80

aaa gag tgg ctg gac cgg gta tta agt cgt ggt ttt gcc agc ggg ccg  288
Lys Glu Trp Leu Asp Arg Val Leu Ser Arg Gly Phe Ala Ser Gly Pro
                85                  90                  95

gga gga aac caa ctg gcg gga aag tac tgg cgt agc gtg att acc acc  336
Gly Gly Asn Gln Leu Ala Gly Lys Tyr Trp Arg Ser Val Ile Thr Thr
            100                 105                 110

ggc gag ccg gaa agt gct tac cgt tat gac gcg ctg aat cgc tac ccg  384
```

```
Gly Glu Pro Glu Ser Ala Tyr Arg Tyr Asp Ala Leu Asn Arg Tyr Pro
        115                 120                 125

atg agc gat gtg ctg cgc ccc ttt gaa ctg gcg gcg ggc atg tgc cgg   432
Met Ser Asp Val Leu Arg Pro Phe Glu Leu Ala Ala Gly Met Cys Arg
    130                 135                 140

atg cat tgg tta agt ccc atc att att tac tgg gcg aga cgg caa agc   480
Met His Trp Leu Ser Pro Ile Ile Ile Tyr Trp Ala Arg Arg Gln Ser
145                 150                 155                 160

gca cag gag ctg gcg agc cac gcc aga gcc tac ggt gac tgg ctg gca   528
Ala Gln Glu Leu Ala Ser His Ala Arg Ala Tyr Gly Asp Trp Leu Ala
                165                 170                 175

aat ccg ctg tct cca gga ggc cgc tga                              555
Asn Pro Leu Ser Pro Gly Gly Arg
            180                 185


<210> 14
<211> 185
<212> PRT
<213> Escherichia coli

<400> 14
Met Met Ser Gln Pro Ala Lys Val Leu Leu Leu Tyr Ala His Pro Glu
  1               5                  10                  15

Ser Gln Asp Ser Val Ala Asn Arg Val Leu Leu Lys Pro Ala Thr Gln
            20                  25                  30

Leu Ser Asn Val Thr Val His Asp Leu Tyr Ala His Tyr Pro Asp Phe
        35                  40                  45

Phe Ile Asp Ile Pro Arg Glu Gln Ala Leu Leu Arg Glu His Glu Val
    50                  55                  60

Ile Val Phe Gln His Pro Leu Tyr Thr Tyr Ser Cys Pro Ala Leu Leu
65                  70                  75                  80

Lys Glu Trp Leu Asp Arg Val Leu Ser Arg Gly Phe Ala Ser Gly Pro
                85                  90                  95

Gly Gly Asn Gln Leu Ala Gly Lys Tyr Trp Arg Ser Val Ile Thr Thr
            100                 105                 110

Gly Glu Pro Glu Ser Ala Tyr Arg Tyr Asp Ala Leu Asn Arg Tyr Pro
        115                 120                 125

Met Ser Asp Val Leu Arg Pro Phe Glu Leu Ala Ala Gly Met Cys Arg
    130                 135                 140

Met His Trp Leu Ser Pro Ile Ile Ile Tyr Trp Ala Arg Arg Gln Ser
145                 150                 155                 160

Ala Gln Glu Leu Ala Ser His Ala Arg Ala Tyr Gly Asp Trp Leu Ala
                165                 170                 175

Asn Pro Leu Ser Pro Gly Gly Arg
            180                 185


<210> 15
<211> 531
<212> DNA
<213> Escherichia coli
```

<220>
<221> CDS
<222> (1)..(531)

<400> 15
atg att ctt ata att tat gcg cat ccg tat ccg cat cat tcc cat gcg 48
Met Ile Leu Ile Ile Tyr Ala His Pro Tyr Pro His His Ser His Ala
1 5 10 15

aat aaa cgg atg ctt gaa cag gca agg acg ctg gaa ggc gtc gaa att 96
Asn Lys Arg Met Leu Glu Gln Ala Arg Thr Leu Glu Gly Val Glu Ile
20 25 30

cgc tct ctt tat caa ctc tat cct gac ttc aat atc gat att gcc gcc 144
Arg Ser Leu Tyr Gln Leu Tyr Pro Asp Phe Asn Ile Asp Ile Ala Ala
35 40 45

gag cag gag gcg ctg tct cgc gcc gat ctg atc gtc tgg cag cat ccg 192
Glu Gln Glu Ala Leu Ser Arg Ala Asp Leu Ile Val Trp Gln His Pro
50 55 60

atg cag tgg tac agc att cct ccg ctc ctc aaa ctt tgg atc gat aaa 240
Met Gln Trp Tyr Ser Ile Pro Pro Leu Leu Lys Leu Trp Ile Asp Lys
65 70 75 80

gtt ttc tcc cac ggc tgg gct tac ggt cat ggc ggc acg gcg ctg cat 288
Val Phe Ser His Gly Trp Ala Tyr Gly His Gly Gly Thr Ala Leu His
85 90 95

ggc aaa cat ttg ctg tgg gcg gtg acg acc ggc ggc ggg gaa agc cat 336
Gly Lys His Leu Leu Trp Ala Val Thr Thr Gly Gly Gly Glu Ser His
100 105 110

ttt gaa att ggt gcg cat ccg ggc ttt gat gtg ctg tcg cag ccg cta 384
Phe Glu Ile Gly Ala His Pro Gly Phe Asp Val Leu Ser Gln Pro Leu
115 120 125

cag gcg acg gca atc tac tgc ggg ctg aac tgg ctg cca ccg ttt gcc 432
Gln Ala Thr Ala Ile Tyr Cys Gly Leu Asn Trp Leu Pro Pro Phe Ala
130 135 140

atg cac tgc acc ttt att tgt gac gac gaa acc ctc gaa ggg cag gcg 480
Met His Cys Thr Phe Ile Cys Asp Asp Glu Thr Leu Glu Gly Gln Ala
145 150 155 160

cgt cac tat aag caa cgt ctg ctg gaa tgg cag gag gcc cat cat gga 528
Arg His Tyr Lys Gln Arg Leu Leu Glu Trp Gln Glu Ala His His Gly
165 170 175

tag 531

<210> 16
<211> 177
<212> PRT
<213> Escherichia coli

<400> 16
Met Ile Leu Ile Ile Tyr Ala His Pro Tyr Pro His His Ser His Ala
1 5 10 15

Asn Lys Arg Met Leu Glu Gln Ala Arg Thr Leu Glu Gly Val Glu Ile
20 25 30

Arg Ser Leu Tyr Gln Leu Tyr Pro Asp Phe Asn Ile Asp Ile Ala Ala
35 40 45

```
Glu Gln Glu Ala Leu Ser Arg Ala Asp Leu Ile Val Trp Gln His Pro
     50                  55                  60

Met Gln Trp Tyr Ser Ile Pro Pro Leu Leu Lys Leu Trp Ile Asp Lys
 65                  70                  75                  80

Val Phe Ser His Gly Trp Ala Tyr Gly His Gly Gly Thr Ala Leu His
                 85                  90                  95

Gly Lys His Leu Leu Trp Ala Val Thr Thr Gly Gly Gly Glu Ser His
            100                 105                 110

Phe Glu Ile Gly Ala His Pro Gly Phe Asp Val Leu Ser Gln Pro Leu
        115                 120                 125

Gln Ala Thr Ala Ile Tyr Cys Gly Leu Asn Trp Leu Pro Pro Phe Ala
    130                 135                 140

Met His Cys Thr Phe Ile Cys Asp Asp Glu Thr Leu Glu Gly Gln Ala
145                 150                 155                 160

Arg His Tyr Lys Gln Arg Leu Leu Glu Trp Gln Glu Ala His His
                165                 170                 175

Gly
```

<210> 17
<211> 222
<212> PRT
<213> Haemophilus influenzae

```
<400> 17
Met Thr Gln Leu Thr Arg Glu Gln Val Leu Glu Leu Phe His Gln Arg
  1                 5                  10                  15

Ser Ser Thr Arg Tyr Tyr Asp Pro Thr Lys Lys Ile Ser Asp Glu Asp
            20                  25                  30

Phe Glu Cys Ile Leu Glu Cys Gly Arg Leu Ser Pro Ser Ser Val Gly
        35                  40                  45

Ser Glu Pro Trp Lys Phe Leu Val Ile Gln Asn Lys Thr Leu Arg Glu
     50                  55                  60

Lys Met Lys Pro Phe Ser Trp Gly Met Ile Asn Gln Leu Asp Asn Cys
 65                  70                  75                  80

Ser His Leu Val Val Ile Leu Ala Lys Lys Asn Ala Arg Tyr Asp Ser
                 85                  90                  95

Gln Gln Gln Ala Ala Leu Thr Lys Tyr Lys Ala Leu Gln Glu Glu Asp
            100                 105                 110

Met Lys Leu Leu Glu Asn Asp Arg Thr Leu Phe Asp Trp Cys Ser Lys
        115                 120                 125

Gln Thr Tyr Ile Ala Leu Ala Asn Met Leu Thr Gly Ala Ser Ala Leu
    130                 135                 140

Gly Ile Asp Ser Cys Pro Ile Glu Gly Phe His Tyr Asp Lys Met Asn
145                 150                 155                 160

Glu Cys Leu Ala Glu Glu Gly Leu Phe Asp Pro Gln Glu Tyr Ala Val
                165                 170                 175

Lys Ser Arg Lys Gly Leu Asp Glu Val Val Lys Trp Val Gly
```

180 185 190

<210> 18
<211> 207
<212> PRT
<213> Thermus aquaticus

<400> 18
Met Glu Ala Thr Leu Pro Val Leu Asp Ala Lys Thr Ala Ala Leu Lys
1 5 10 15

Arg Arg Ser Ile Arg Arg Tyr Arg Lys Asp Pro Val Pro Glu Gly Leu
20 25 30

Leu Arg Glu Ile Leu Glu Ala Ala Leu Arg Ala Pro Ser Ala Trp Asn
35 40 45

Leu Gln Pro Trp Arg Ile Val Val Val Arg Asp Pro Ala Thr Lys Arg
50 55 60

Ala Leu Arg Glu Ala Ala Phe Gly Gln Ala His Val Glu Glu Ala Pro
65 70 75 80

Val Val Leu Val Leu Tyr Ala Asp Leu Glu Asp Ala Leu Ala His Leu
85 90 95

Gln Lys Gln Ala Ile Gln Arg Ala Phe Ala Ala Met Gly Gln Glu Ala
100 105 110

Arg Lys Ala Trp Ala Ser Gly Gln Ser Tyr Ile Leu Leu Gly Tyr Leu
115 120 125

Leu Leu Leu Leu Glu Ala Tyr Gly Leu Gly Ser Val Pro Met Leu Gly
130 135 140

Phe Asp Pro Glu Arg Val Arg Ala Ile Leu Gly Leu Pro Ser Arg Ala
145 150 155 160

Ala Ile Pro Ala Leu Val Ala Leu Gly Tyr Pro Ala Glu Glu Gly Tyr
165 170 175

Pro Ser His Arg Leu Pro Leu Glu Arg 0 Val Val Leu Trp Arg
180 185 0 190

<210> 19
<211> 212
<212> PRT
<213> Synechocystis PCC6803

<400> 19
Met Asp Thr Phe Asp Ala Ile Tyr Gln Arg Arg Ser Val Lys His Phe
1 5 10 15

Asp Pro Asp His Arg Leu Thr Ala Glu Glu Glu Arg Lys Leu His Glu
20 25 30

Ala Ala Ile Gln Ala Pro Thr Ser Phe Asn Ile Gln Leu Trp Arg Phe
35 40 45

Leu Ile Ile Arg Asp Pro Gln Leu Arg Gln Thr Ile Arg Glu Lys Tyr
50 55 60

Gly Asn Gln Ala Gln Met Thr Asp Ala Ser Leu Leu Ile Leu Val Ala
65 70 75 80

```
Ala Asp Val Asn Ala Trp Asp Lys Asp Pro Ala Arg Tyr Trp Arg Asn
                  85                  90                  95

Phe Tyr Gly Gly Lys Pro Gln Leu Gln Arg Asp Glu Ala Gln Arg Ser
             100                 105                 110

Ile Gly Met Ala Met Gln Asn Leu Met Leu Ala Ala Lys Ala Met Gly
         115                 120                 125

Tyr Asp Ser Cys Pro Met Ile Gly Phe Asp Leu Gln Lys Val Ala Glu
     130                 135                 140

Leu Val Lys Leu Pro Ala Asp Tyr Ala Ile Gly Pro Met Val Ala Ile
145                 150                 155                 160

Gly Lys Arg Thr Glu Asp Ala Pro Gly Lys Arg Arg Ser Asn Ser Pro
                 165                 170                 175

Cys Leu Ala Ile
             180
```

<210> 20
<211> 172
<212> PRT
<213> Archaeoglobus fulgidus

```
<400> 20
Met Glu Cys Leu Asp Leu Leu Phe Arg Arg Val Ser Ile Arg Lys Phe
  1                 5                  10                  15

Thr Gln Asp Asp Val Asp Asp Glu Ile Leu Met Lys Ile Leu Glu Ala
             20                  25                  30

Gly Asn Ala Ala Pro Ser Ala Gly Asn Leu Gln Ala Arg Asp Phe Val
         35                  40                  45

Val Ile Arg Asn Pro Glu Thr Lys Lys Arg Leu Ala Met Ala Ala Leu
     50                  55                  60

Lys Gln Met Phe Ile Ala Glu Ala Pro Val Val Ile Val Val Cys Ala
 65                  70                  75                  80

Asn Tyr Pro Arg Ser Met Arg Val Tyr Gly Glu Arg Gly Arg Leu Tyr
                 85                  90                  95

Ala Glu Gln Asp Ala Thr Ala Ala Ile Glu Asn Ile Leu Leu Ala Val
             100                 105                 110

Thr Ala Leu Asn Leu Gly Ala Val Trp Val Gly Ala Phe Asp Glu Glu
         115                 120                 125

Gln Val Ser Glu Ile Leu Glu Leu Pro Glu Tyr Val Arg Pro Met Ala
     130                 135                 140

Ile Ile Pro Ile Gly His Pro Ala Glu Asn Pro Ser Pro Arg Asn Arg
145                 150                 155                 160

Tyr Pro Val Ser Met Leu Thr His Phe Glu Lys Trp
                 165                 170
```

<210> 21
<211> 174
<212> PRT
<213> Archaeoglobus fulgidus

<400> 21
Met Glu Glu Cys Leu Lys Met Ile Tyr Thr Arg Arg Ser Ile Arg Val
1 5 10 15

Tyr Ser Asp Arg Gln Ile Ser Asp Glu Asp Ile Glu Lys Ile Leu Lys
20 25 30

Ala Ala Met Leu Ala Pro Ser Ala Gly Asn Glu Gln Pro Trp His Phe
35 40 45

Ile Val Val Arg Asp Arg Glu Met Leu Lys Lys Met Ser Glu Ala Phe
50 55 60

Thr Phe Gly Gln Met Leu Pro Asn Ala Ser Ala Ala Ile Val Val Cys
65 70 75 80

Ala Asp Pro Lys Leu Ser Lys Tyr Pro Tyr Asp Met Trp Val Gln Asp
85 90 95

Cys Ser Ala Ala Thr Glu Asn Ile Leu Leu Ala Ala Arg Cys Leu Gly
100 105 110

Ile Gly Ser Val Trp Leu Gly Val Tyr Pro Arg Glu Glu Arg Met Lys
115 120 125

Ala Leu Arg Glu Leu Leu Gly Ile Pro Glu Asn Ile Val Val Phe Ser
130 135 140

Val Val Ser Leu Gly Tyr Pro Lys Asp Glu Lys Asp Phe Tyr Glu Ala
145 150 155 160

Asp Asp Arg Phe Asn Pro Asp Arg Ile His Arg Glu Lys Trp
165 170

<210> 22
<211> 606
<212> DNA
<213> Campylobacter jejuni

<220>
<221> CDS
<222> (1)..(606)

<400> 22
atg aaa aaa gaa ctt gaa att ttt agc aca aga tat tct tgt aga aat 48
Met Lys Lys Glu Leu Glu Ile Phe Ser Thr Arg Tyr Ser Cys Arg Asn
1 5 10 15

ttt aaa aat gaa aaa ctc aaa aaa gag gat tta aat tct att tta gaa 96
Phe Lys Asn Glu Lys Leu Lys Lys Glu Asp Leu Asn Ser Ile Leu Glu
20 25 30

ata gca aga tta agc ccc agt tcc ttg gga ctg gaa cct tgg aaa ttt 144
Ile Ala Arg Leu Ser Pro Ser Ser Leu Gly Leu Glu Pro Trp Lys Phe
35 40 45

ata gta gtg caa gat gag aaa aga aaa gaa gaa ctt tct aaa att tgc 192
Ile Val Val Gln Asp Glu Lys Arg Lys Glu Glu Leu Ser Lys Ile Cys
50 55 60

aat caa caa aaa cat gta aaa gat tgt gct gca tta att ata atc att 240
Asn Gln Gln Lys His Val Lys Asp Cys Ala Ala Leu Ile Ile Ile Ile
65 70 75 80

tca aga ctt gat ttt ttg gat tat ttt gaa gaa aaa ctt aga aaa aga 288
Ser Arg Leu Asp Phe Leu Asp Tyr Phe Glu Glu Lys Leu Arg Lys Arg

```
                    85                  90                  95

gat atg agt gaa aca gaa atg caa aaa cgc tta gat act tat atg cct   336
Asp Met Ser Glu Thr Glu Met Gln Lys Arg Leu Asp Thr Tyr Met Pro
            100                 105                 110

ttt tta aaa tct cta aat caa gaa caa aaa ata tct tat gca aga gaa   384
Phe Leu Lys Ser Leu Asn Gln Glu Gln Lys Ile Ser Tyr Ala Arg Glu
        115                 120                 125

caa gct cat ata gct cta gct agc ata ctt tac agt gct aat gct tta   432
Gln Ala His Ile Ala Leu Ala Ser Ile Leu Tyr Ser Ala Asn Ala Leu
    130                 135                 140

aat ata gca agc tgc act ata ggt ggt ttt gat aaa gaa aag ctt gat   480
Asn Ile Ala Ser Cys Thr Ile Gly Gly Phe Asp Lys Glu Lys Leu Asp
145                 150                 155                 160

tct tat tta tca ctt gat att caa aaa gaa aga tca agt ttg gtg gtg   528
Ser Tyr Leu Ser Leu Asp Ile Gln Lys Glu Arg Ser Ser Leu Val Val
                165                 170                 175

gct tta gga tat tgc aac gat aaa aaa aat cct caa aaa aat cgt ttt   576
Ala Leu Gly Tyr Cys Asn Asp Lys Lys Asn Pro Gln Lys Asn Arg Phe
            180                 185                 190

agt ttt gat gaa gtt gta aaa ttt att taa                          606
Ser Phe Asp Glu Val Val Lys Phe Ile
        195                 200
```

<210> 23
<211> 202
<212> PRT
<213> Campylobacter jejuni

<400> 23

```
Met Lys Lys Glu Leu Glu Ile Phe Ser Thr Arg Tyr Ser Cys Arg Asn
  1               5                  10                  15

Phe Lys Asn Glu Lys Leu Lys Lys Glu Asp Leu Asn Ser Ile Leu Glu
            20                  25                  30

Ile Ala Arg Leu Ser Pro Ser Ser Leu Gly Leu Glu Pro Trp Lys Phe
        35                  40                  45

Ile Val Val Gln Asp Glu Lys Arg Lys Glu Glu Leu Ser Lys Ile Cys
    50                  55                  60

Asn Gln Gln Lys His Val Lys Asp Cys Ala Ala Leu Ile Ile Ile Ile
65                  70                  75                  80

Ser Arg Leu Asp Phe Leu Asp Tyr Phe Glu Glu Lys Leu Arg Lys Arg
                85                  90                  95

Asp Met Ser Glu Thr Glu Met Gln Lys Arg Leu Asp Thr Tyr Met Pro
            100                 105                 110

Phe Leu Lys Ser Leu Asn Gln Glu Gln Lys Ile Ser Tyr Ala Arg Glu
        115                 120                 125

Gln Ala His Ile Ala Leu Ala Ser Ile Leu Tyr Ser Ala Asn Ala Leu
    130                 135                 140

Asn Ile Ala Ser Cys Thr Ile Gly Gly Phe Asp Lys Glu Lys Leu Asp
145                 150                 155                 160
```

```
Ser Tyr Leu Ser Leu Asp Ile Gln Lys Glu Arg Ser Ser Leu Val Val
                165                 170                 175

Ala Leu Gly Tyr Cys Asn Asp Lys Lys Asn Pro Gln Lys Asn Arg Phe
            180                 185                 190

Ser Phe Asp Glu Val Val Lys Phe Ile
        195                 200
```

<210> 24
<211> 522
<212> DNA
<213> Porphyromonas gingivalis

<220>
<221> CDS
<222> (1)..(522)

<400> 24

```
atg aaa aaa acg ctc gta ata gtc gtt cac ccc gat ttg acc aaa tcc    48
Met Lys Lys Thr Leu Val Ile Val Val His Pro Asp Leu Thr Lys Ser
  1               5                  10                  15

gtt atc aac aag gct tgg gcc aaa gcc atc gaa ggt gca gcc act atc    96
Val Ile Asn Lys Ala Trp Ala Lys Ala Ile Glu Gly Ala Ala Thr Ile
            20                  25                  30

cac cat ctc tac gaa cag tat ccg aac gga caa atc gat cta gca cat   144
His His Leu Tyr Glu Gln Tyr Pro Asn Gly Gln Ile Asp Leu Ala His
        35                  40                  45

gaa caa gcc ctg ctg gag gct cat gac cgc atc gtc ttc caa ttc ccc   192
Glu Gln Ala Leu Leu Glu Ala His Asp Arg Ile Val Phe Gln Phe Pro
    50                  55                  60

ctc tat tgg tat gca gct ccc tat ctg ctg aag aag tgg atg gac gag   240
Leu Tyr Trp Tyr Ala Ala Pro Tyr Leu Leu Lys Lys Trp Met Asp Glu
65                  70                  75                  80

gtc ttt act gag ggc tgg gcc tat ggt gcc ggt gga gac aag atg gag   288
Val Phe Thr Glu Gly Trp Ala Tyr Gly Ala Gly Gly Asp Lys Met Glu
                85                  90                  95

ggt aaa gaa atc tgt gca gca gtc tcc tgc gga tca ccc aaa tca gct   336
Gly Lys Glu Ile Cys Ala Ala Val Ser Cys Gly Ser Pro Lys Ser Ala
            100                 105                 110

ttt gcc gaa gga gca cag caa tgc cac acg ctg cga agc tac ttg aat   384
Phe Ala Glu Gly Ala Gln Gln Cys His Thr Leu Arg Ser Tyr Leu Asn
        115                 120                 125

gta ttc gac ggg ata gct gct ttc ctg cgc gct cga ttc acc ggc tac   432
Val Phe Asp Gly Ile Ala Ala Phe Leu Arg Ala Arg Phe Thr Gly Tyr
    130                 135                 140

cat gcc tgc tac gat tcc tac aat cct cgc ctg ccg gaa atg ctg ccg   480
His Ala Cys Tyr Asp Ser Tyr Asn Pro Arg Leu Pro Glu Met Leu Pro
145                 150                 155                 160

gcc aac tgc gaa gcc tat ctc cgc ttt atc aaa gga gaa tga           522
Ala Asn Cys Glu Ala Tyr Leu Arg Phe Ile Lys Gly Glu
                165                 170
```

<210> 25
<211> 174

<212> PRT
<213> Porphyromonas gingivalis

<400> 25
Met Lys Lys Thr Leu Val Ile Val Val His Pro Asp Leu Thr Lys Ser
1 5 10 15

Val Ile Asn Lys Ala Trp Ala Lys Ala Ile Glu Gly Ala Ala Thr Ile
20 25 30

His His Leu Tyr Glu Gln Tyr Pro Asn Gly Gln Ile Asp Leu Ala His
35 40 45

Glu Gln Ala Leu Leu Glu Ala His Asp Arg Ile Val Phe Gln Phe Pro
50 55 60

Leu Tyr Trp Tyr Ala Ala Pro Tyr Leu Leu Lys Lys Trp Met Asp Glu
65 70 75 80

Val Phe Thr Glu Gly Trp Ala Tyr Gly Ala Gly Gly Asp Lys Met Glu
85 90 95

Gly Lys Glu Ile Cys Ala Ala Val Ser Cys Gly Ser Pro Lys Ser Ala
100 105 110

Phe Ala Glu Gly Ala Gln Gln Cys His Thr Leu Arg Ser Tyr Leu Asn
115 120 125

Val Phe Asp Gly Ile Ala Ala Phe Leu Arg Ala Arg Phe Thr Gly Tyr
130 135 140

His Ala Cys Tyr Asp Ser Tyr Asn Pro Arg Leu Pro Glu Met Leu Pro
145 150 155 160

Ala Asn Cys Glu Ala Tyr Leu Arg Phe Ile Lys Gly Glu
165 170

<210> 26
<211> 552
<212> DNA
<213> Yersinia pestis

<220>
<221> CDS
<222> (1)..(552)

<400> 26
atg atg ttg cag ccg ccg aag gtt ttg ctg ctg tat gcc cat ccg gaa 48
Met Met Leu Gln Pro Pro Lys Val Leu Leu Leu Tyr Ala His Pro Glu
1 5 10 15

tca cag gac tcg gtc gct aac cgg gtt tta ctg caa ccg gta cag cag 96
Ser Gln Asp Ser Val Ala Asn Arg Val Leu Leu Gln Pro Val Gln Gln
20 25 30

tta gaa cat gtc act gtg cac gat ctt tat gca cat tat ccg gat ttc 144
Leu Glu His Val Thr Val His Asp Leu Tyr Ala His Tyr Pro Asp Phe
35 40 45

ttt att gat att cat cat gag cag caa ttg cta cgt gat cat caa gtt 192
Phe Ile Asp Ile His His Glu Gln Gln Leu Leu Arg Asp His Gln Val
50 55 60

att gta ttt caa cat cct tta tat act tac agt tgc cct gca tta ctg 240
Ile Val Phe Gln His Pro Leu Tyr Thr Tyr Ser Cys Pro Ala Leu Leu
65 70 75 80

```
aaa gag tgg ttg gat cgg gta ctg gca cgt ggt ttc gcc aat ggc gtt   288
Lys Glu Trp Leu Asp Arg Val Leu Ala Arg Gly Phe Ala Asn Gly Val
                85                  90                  95

ggc ggc cat gca ctg acg gga aag cac tgg cgc tcg gtg att acc acc   336
Gly Gly His Ala Leu Thr Gly Lys His Trp Arg Ser Val Ile Thr Thr
            100                 105                 110

ggt gag cag gag gga act tac cgt att ggg gga tat aac cgt tac cca   384
Gly Glu Gln Glu Gly Thr Tyr Arg Ile Gly Gly Tyr Asn Arg Tyr Pro
        115                 120                 125

atg gaa gac att ctg cgt cct ttc gaa ttg acg gcg gct atg tgc cat   432
Met Glu Asp Ile Leu Arg Pro Phe Glu Leu Thr Ala Ala Met Cys His
    130                 135                 140

atg cat tgg att aat ccg atg att att tac tgg gcc aga cgc caa aag   480
Met His Trp Ile Asn Pro Met Ile Ile Tyr Trp Ala Arg Arg Gln Lys
145                 150                 155                 160

ccg gaa aca ctc gcc agt cac gca caa gct tat gtg caa tgg ctg cag   528
Pro Glu Thr Leu Ala Ser His Ala Gln Ala Tyr Val Gln Trp Leu Gln
                165                 170                 175

tca ccg ctc acg aga gga ctc tga                                  552
Ser Pro Leu Thr Arg Gly Leu
            180
```

<210> 27
<211> 184
<212> PRT
<213> Yersinia pestis

<400> 27

```
Met Met Leu Gln Pro Pro Lys Val Leu Leu Leu Tyr Ala His Pro Glu
  1               5                  10                  15

Ser Gln Asp Ser Val Ala Asn Arg Val Leu Leu Gln Pro Val Gln Gln
             20                  25                  30

Leu Glu His Val Thr Val His Asp Leu Tyr Ala His Tyr Pro Asp Phe
         35                  40                  45

Phe Ile Asp Ile His His Glu Gln Gln Leu Leu Arg Asp His Gln Val
     50                  55                  60

Ile Val Phe Gln His Pro Leu Tyr Thr Tyr Ser Cys Pro Ala Leu Leu
 65                  70                  75                  80

Lys Glu Trp Leu Asp Arg Val Leu Ala Arg Gly Phe Ala Asn Gly Val
                85                  90                  95

Gly Gly His Ala Leu Thr Gly Lys His Trp Arg Ser Val Ile Thr Thr
            100                 105                 110

Gly Glu Gln Glu Gly Thr Tyr Arg Ile Gly Gly Tyr Asn Arg Tyr Pro
        115                 120                 125

Met Glu Asp Ile Leu Arg Pro Phe Glu Leu Thr Ala Ala Met Cys His
    130                 135                 140

Met His Trp Ile Asn Pro Met Ile Ile Tyr Trp Ala Arg Arg Gln Lys
145                 150                 155                 160

Pro Glu Thr Leu Ala Ser His Ala Gln Ala Tyr Val Gln Trp Leu Gln
                165                 170                 175
```

Ser Pro Leu Thr Arg Gly Leu
180

<210> 28
<211> 633
<212> DNA
<213> Helicobacter pylori

<400> 28
atgaaatttt tggatcaaga aaaaagaaga caattgctaa acgagcgcca ttcttgcaag 60

atgttcgaca gccattatga gttttctagt gaagaattag aagaaatcgc tgaaatcgct 120

aggctatcgc caagctctta caacacgcag ccatggcatt ttgtgatggt tactaataag 180

gatttaaaaa aacaaattgc agcgcacagc tattttaatg aagaaatgat taaaagcgct 240

tcagcgttaa tggtggtatg ctctttaaaa cccagcgagt tgttacccac tggccactac 300

atgcaaaacc tttacccgga gtcttataag gttagagtga tcccctcttt tgctcaaatg 360

cttggcgtga gattcaacca cagcatgcaa aaattagaaa gctatatttt ggagcaatgc 420

tatatcgctg tggggcaaat ttgcatgggc gtgagcttaa tgggattgga tagttgcatt 480

attggaggct ttgatccttt aaaagtgggc gaagttttag aagagcgtat caataaacct 540

aaaatcgcat gcttgatcgc tttgggcaag agggtggcag aagcgagcca aaaatcaaga 600

aaatcaaaag ttgatgccat tacttggttg tga 633

<210> 29
<211> 210
<212> PRT
<213> Helicobacter pylori

<400> 29
Met Lys Phe Leu Asp Gln Glu Lys Arg Arg Gln Leu Leu Asn Glu Arg
1 5 10 15

His Ser Cys Lys Met Phe Asp Ser His Tyr Glu Phe Ser Ser Glu Glu
20 25 30

Leu Glu Glu Ile Ala Glu Ile Ala Arg Leu Ser Pro Ser Ser Tyr Asn
35 40 45

Thr Gln Pro Trp His Phe Val Met Val Thr Asn Lys Asp Leu Lys Lys
50 55 60

Gln Ile Ala Ala His Ser Tyr Phe Asn Glu Glu Met Ile Lys Ser Ala
65 70 75 80

Ser Ala Leu Met Val Val Cys Ser Leu Lys Pro Ser Glu Leu Leu Pro
85 90 95

Thr Gly His Tyr Met Gln Asn Leu Tyr Pro Glu Ser Tyr Lys Val Arg
100 105 110

Val Ile Pro Ser Phe Ala Gln Met Leu Gly Val Arg Phe Asn His Ser
115 120 125

Met Gln Lys Leu Glu Ser Tyr Ile Leu Glu Gln Cys Tyr Ile Ala Val
130 135 140

Gly Gln Ile Cys Met Gly Val Ser Leu Met Gly Leu Asp Ser Cys Ile

```
145                 150                 155                 160

Ile Gly Gly Phe Asp Pro Leu Lys Val Gly Glu Val Leu Glu Glu Arg
                   165                 170                 175

Ile Asn Lys Pro Lys Ile Ala Cys Leu Ile Ala Leu Gly Lys Arg Val
           180                 185                 190

Ala Glu Ala Ser Gln Lys Ser Arg Lys Ser Lys Val Asp Ala Ile Thr
       195                 200                 205

Trp Leu
   210
```

# INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCT/GB 00/00431

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N9/02 C12N15/52 A61K35/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, EPO-Internal, STRAND, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | EP 0 540 263 A (CANCER RES CAMPAIGN TECH) 5 May 1993 (1993-05-05) cited in the application the whole document | 1-3,5-28 |
| X | WO 95 12678 A (CONNORS THOMAS ;KNOX RICHARD (GB); SHERWOOD ROGER (GB); CANCER RES) 11 May 1995 (1995-05-11) the whole document especially figure 6, examples 1-4 | 1-3,5-28 |
| X | DE 42 21 830 A (BIOTECHNOLOG FORSCHUNG GMBH) 28 January 1993 (1993-01-28) the whole document | 1-3,5-28 |

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 13 July 2000 | 25/07/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Panzica, G |

1

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/GB 00/00431

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | EP 0 547 876 A (CHISSO CORP)<br>23 June 1993 (1993-06-23)<br>abstract<br>claim 4; figure 4 | 1-3,5-28 |
| X | ANTELMANN H. ET AL.: "First step from a two-dimensional protein index towards a response-regulation map for Bacillus subtilis"<br>ELECTROPHORESIS,<br>vol. 18, no. 8, 1997, pages 1451-1463,<br>XP000923464<br>the whole document | 1-3,5-28 |
| X | WO 98 57662 A (BURKE PHILIP JOHN ;ENZACTA R & D LTD (GB); KNOX RICHARD JOHN (GB))<br>23 December 1998 (1998-12-23)<br>abstract<br>figure 6; example 1 | 1-3,5-28 |

Form PCT/ISA/210 (continuation of second sheet) (July 1992)